

University of Warwick institutional repository: <http://go.warwick.ac.uk/wrap>

A Thesis Submitted for the Degree of PhD at the University of Warwick

<http://go.warwick.ac.uk/wrap/36361>

This thesis is made available online and is protected by original copyright.

Please scroll down to view the document itself.

Please refer to the repository record for this item for information to help you to cite it. Our policy information is available from the repository home page.



**PERFORMANCE MEASUREMENT AND IMPROVEMENT
IN THE MANAGEMENT OF BANK BRANCH NETWORKS
USING DATA ENVELOPMENT ANALYSIS**

by

Ana Maria Cunha Ribeiro dos Santos Ponces Camanho

A thesis submitted for the degree of Doctor of Philosophy

Operational Research and Systems
Warwick Business School
University of Warwick

September 1999

TABLE OF CONTENTS

TABLE OF CONTENTS	i
LIST OF FIGURES.....	viii
LIST OF TABLES.....	xi
ACKNOWLEDGMENTS.....	xiv
DECLARATION	xv
ABSTRACT	xvi
GLOSSARY	xvii
 CHAPTER 1 INTRODUCTION.....	 1
1.1 INTRODUCTION	1
1.2 MOTIVATION AND RESEARCH OBJECTIVES	4
1.3 THESIS STRUCTURE	7
 CHAPTER 2 THE ASSESSMENT OF PERFORMANCE	 9
2.1 INTRODUCTION	9
2.2 HISTORICAL EVOLUTION OF FRONTIER ANALYSIS	11
2.3 OVERVIEW OF FRONTIER ANALYSIS METHODS	17
2.3.1 <i>Parametric deterministic frontiers</i>	17
2.3.2 <i>Parametric stochastic frontiers</i>	18
2.3.3 <i>Non-parametric deterministic frontiers</i>	20
2.3.4 <i>Non-parametric stochastic frontiers</i>	21
2.4 DATA ENVELOPMENT ANALYSIS	22

2.4.1	<i>Production possibility set and efficient frontiers</i>	22
2.4.2	<i>The approach</i>	27
2.4.3	<i>The DEA models</i>	28
2.4.3.1	The ratio model	29
2.4.3.2	The DEA model with constant returns to scale	30
2.4.3.3	The DEA model with variable returns to scale.....	34
2.4.4	<i>Scale efficiency and most productive scale size</i>	40
2.4.5	<i>Economic efficiency</i>	43
2.4.6	<i>Recent developments</i>	47
2.4.6.1	Need of a performance measure and goals of the analysis	48
2.4.6.2	Selection of the DEA method.....	49
2.4.6.3	Selection of time period for assessment	50
2.4.6.4	Selection of DMUs.....	51
2.4.6.5	Selection of input-output variables.....	51
2.4.6.6	Selection of the DEA model(s).....	53
2.4.6.7	Weights restrictions.....	54
2.4.6.8	Efficiency results.....	54
2.4.6.9	Efficiency improvement: Peers and Targets.....	55
2.4.6.10	Conveying results to decision makers	55
2.4.6.11	Implementation of the results.....	56
2.5	SUMMARY AND CONCLUSIONS	56

CHAPTER 3 THE ANALYSIS OF EFFICIENCY IN THE BANKING SECTOR: AN

OVERVIEW OF PAST RESEARCH AND FUTURE DIRECTIONS57

3.1	INTRODUCTION	57
3.2	ALTERNATIVE METHODS FOR THE ASSESSMENT OF PERFORMANCE IN BANKING	59
3.2.1	<i>Ratio analysis</i>	59
3.2.2	<i>Profitability</i>	61
3.2.3	<i>Frontier methods</i>	63
3.3	THE INPUTS AND OUTPUTS OF BANKING	66
3.3.1	<i>Production Approach</i>	67
3.3.2	<i>Intermediation Approach</i>	69

3.3.3	<i>Asset Approach</i>	71
3.3.4	<i>User Cost Approach</i>	72
3.3.5	<i>Value-Added Approach</i>	72
3.3.6	<i>Discussion</i>	73
3.4	THE MEASUREMENT OF EFFICIENCY IN BANKING	74
3.4.1	<i>The impact of frontier specification on efficiency estimates</i>	76
3.4.2	<i>Sources of banks inefficiency</i>	77
3.4.3	<i>Definition of inputs and outputs and sensitivity of the efficiency measures to variable specification</i>	79
3.4.4	<i>Orientation of the assessment</i>	82
3.4.5	<i>The effect of deregulation on banking efficiency</i>	83
3.4.6	<i>Prediction of financial institution failure</i>	84
3.4.7	<i>Organisational form and ownership impact on efficiency</i>	84
3.4.8	<i>Mergers impact on efficiency</i>	85
3.4.9	<i>Inter-country comparisons</i>	86
3.4.10	<i>Improving the performance of branches within a financial institution</i>	87
3.5	SUMMARY AND CONCLUSIONS	90
 CHAPTER 4 INTRODUCTION TO THE PORTUGUESE BANKING SECTOR AND THE		
	CASE STUDY OF THE PORTUGUESE BANK	93
4.1	INTRODUCTION	93
4.2	OVERVIEW OF THE PORTUGUESE BANKING SECTOR.....	94
4.2.1	<i>Introduction</i>	94
4.2.2	<i>Structure of the banking sector</i>	97
4.2.3	<i>Market share of the institutions</i>	100
4.2.4	<i>Competitive conditions and financial margins</i>	102
4.3	THE CASE STUDY: INTRODUCTION TO THE PORTUGUESE BANK.....	104
4.3.1	<i>Description of the bank</i>	105
4.3.2	<i>Performance measurement methods at the Bank</i>	108
4.3.2.1	<i>Earning Analysis System</i>	109
4.3.2.2	<i>System of Incentives and Motivation</i>	111

4.3.3	<i>Motivation to use the DEA method</i>	111
4.4	SUMMARY AND CONCLUSIONS	112
 CHAPTER 5 EFFICIENCY AND PROFITABILITY MEASUREMENT IN BANK BRANCH NETWORKS		
		114
5.1	INTRODUCTION	114
5.2	FRAMEWORK FOR THE EFFICIENCY ASSESSMENT	118
5.3	BRANCHES' ACTIVITY MODELS	124
5.3.1	<i>Production approach</i>	126
5.3.2	<i>Intermediation approach</i>	129
5.4	EMPIRICAL RESULTS OF THE EFFICIENCY ASSESSMENT	133
5.4.1	<i>Selection of appropriate DEA models regarding returns to scale</i>	133
5.4.2	<i>Intermediation efficiency analysis</i>	137
5.4.2.1	Cost-effectiveness measure	137
5.4.2.2	Technical efficiency	139
5.4.2.3	Input allocative efficiency	139
5.4.2.4	Output mix efficiency.....	140
5.4.2.5	Comparison of efficiency results with different output sets	140
5.4.3	<i>Production efficiency analysis</i>	143
5.4.3.1	Cost efficiency measure	143
5.4.3.2	Pure technical efficiency	144
5.4.3.3	Scale efficiency	145
5.4.3.4	Input allocative efficiency	147
5.5	MANAGERIAL IMPLICATIONS OF THE EFFICIENCY ASSESSMENT	147
5.5.1	<i>Comparative analysis of production and intermediation efficiency</i>	147
5.5.2	<i>Implications of bank mergers in terms of branches' efficiency level</i>	150
5.6	EFFICIENCY AND PROFITABILITY.....	152
5.7	SUMMARY AND CONCLUSIONS	157

CHAPTER 6	IMPROVING OPERATIONAL ACTIVITY: PURE TECHNICAL AND SCALE EFFICIENCY ISSUES	160
6.1	INTRODUCTION	160
6.2	INPUT AND OUTPUT MEASURES	161
6.3	EFFECT OF SCALE SIZE ON BRANCHES' OPERATIONAL ACTIVITY	164
6.3.1	<i>Definition of branch scale size.....</i>	<i>165</i>
6.3.2	<i>Input versus output oriented assessments.....</i>	<i>166</i>
6.3.3	<i>The components of the operational efficiency measure.....</i>	<i>169</i>
6.3.4	<i>Returns to scale characterisation.....</i>	<i>172</i>
6.4	TARGET SETTING	175
6.4.1	<i>Targets with selection of appropriate benchmarks.....</i>	<i>175</i>
6.4.1.1	Motivation.....	175
6.4.1.2	DEA model with restricted peers	176
6.4.1.3	Empirical results.....	179
6.4.2	<i>Most productive scale size targets.....</i>	<i>180</i>
6.4.2.1	Motivation.....	180
6.4.2.2	Method to derive the MPSS target closest to the DMUs' PTE scale size.....	182
6.4.2.3	Empirical results.....	185
6.5	SUMMARY AND CONCLUSIONS	186
CHAPTER 7	COST EFFICIENCY UNDER DIFFERENT PRICE SCENARIOS	188
7.1	INTRODUCTION	188
7.2	COST EFFICIENCY WITH FIXED PRICES	190
7.2.1	<i>Prices known at the DMU level.....</i>	<i>190</i>
7.2.2	<i>Prices unknown at the DMU level.....</i>	<i>193</i>
7.2.2.1	Graphical illustration of efficiency measurement in the 'weights' space	194
7.2.2.2	The optimistic perspective of CE measurement	198
7.2.2.3	The pessimistic perspective of CE measurement	201
7.3	COST EFFICIENCY WITH VARIABLE PRICES	207
7.3.1	<i>Introduction.....</i>	<i>207</i>
7.3.2	<i>The economic efficiency components.....</i>	<i>208</i>

7.3.3	<i>Target setting with emphasis on keeping the current input mix.....</i>	213
7.3.3.1	Illustration of the target setting method.....	216
7.4	ILLUSTRATION OF THE METHODS DEVELOPED WITH THE ASSESSMENT OF BANK BRANCHES	220
7.4.1	<i>Cost efficiency with fixed prices</i>	222
7.4.1.1	Prices known at the DMU level.....	222
7.4.1.2	Prices unknown at the DMU level.....	223
7.4.2	<i>Cost efficiency with variable prices.....</i>	225
7.4.2.1	Economic efficiency components.....	225
7.4.2.2	Target setting with emphasis on keeping the current input mix	226
7.4.3	<i>Comparison of CE measures</i>	227
7.4.3.1	Overall network level.....	227
7.4.3.2	Individual branch level.....	228
7.5	SUMMARY AND CONCLUSIONS	230
CHAPTER 8	ASSESSMENT OF BANK BRANCHES' PERFORMANCE AT THE REGIONAL LEVEL: COMPARISON OF REGIONAL EFFICIENCY SPREAD AND FRONTIER PRODUCTIVITY.....	233
8.1	INTRODUCTION	233
8.2	REVIEW OF THE 'PROGRAM EFFICIENCY' METHOD	235
8.3	THE MALMQUIST INDEX	237
8.4	NEW METHOD FOR COMPARING THE PERFORMANCE OF DMUS OPERATING UNDER DIFFERENT CONDITIONS.....	244
8.5	COMPUTATION OF DISTANCE FUNCTIONS USING THE DEA APPROACH.....	251
8.6	DESCRIPTION OF THE EMPIRICAL DATA	253
8.7	EMPIRICAL RESULTS	254
8.8	SUMMARY AND CONCLUSIONS	261
CHAPTER 9	SUMMARY, CONCLUSIONS AND DIRECTIONS FOR FURTHER RESEARCH..	263
9.1	SUMMARY AND CONCLUSIONS	263
9.2	DIRECTIONS FOR FURTHER RESEARCH	275

APPENDIX TO CHAPTER 3	278
SUMMARY OF NON-PARAMETRIC STUDIES ON BANKING EFFICIENCY	278
 APPENDIX TO CHAPTER 7	 298
APPENDIX 7A - ILLUSTRATION OF EFFICIENCY MEASUREMENT IN THE WEIGHTS SPACE.....	298
APPENDIX 7B - ILLUSTRATION OF THE WEIGHTS RESTRICTIONS FOR RELATIVE PRICES USED IN THE BANK BRANCHES' ASSESSMENT	 304
 REFERENCES	 306

LIST OF FIGURES

Chapter 2

Figure 2.1 - <i>The production process</i>	11
Figure 2.2 - <i>Efficiency measurement</i>	14
Figure 2.3 - <i>Classification of production frontiers</i>	16
Figure 2.4 - <i>Farrell's versus Koopmans' notions of efficiency</i>	26
Figure 2.5 - <i>CRS versus VRS frontiers</i>	35
Figure 2.6 - <i>Scale efficiency notion</i>	41
Figure 2.7 - <i>Efficiency components</i>	47
Figure 2.8 - <i>The DEA application procedure</i>	48

Chapter 4

Figure 4.1 - <i>Number of banks operating in Portugal between 1984 and 1997 (source: Bank of Portugal, 1997 Annual Report)</i>	98
Figure 4.2 - <i>Evolution of the number of bank branches operating in Portugal (source: Official report of the Portuguese Banking Association, June 1997 and Bank of Portugal, 1997 Annual Report).</i>	99
Figure 4.3 - <i>Market share of state owned banks (source: Bank of Portugal, 1997 Annual Report)</i>	100
Figure 4.4 - <i>Market share of the five major banking groups (source: Bank of Portugal, 1997 Annual Report)</i>	101
Figure 4.5 - <i>Income and costs in the Portuguese banking sector (source: Bank of Portugal, 1997 Annual Report)</i>	102
Figure 4.6 - <i>International comparison of financial margins (source: OECD, bank profitability, Paris 1997)</i>	103

Figure 4.7 – <i>Breakdown of the banking product (source: Bank of Portugal, 1997 Annual Report)</i>	104
Figure 4.8 – <i>Branch network expansion</i>	106
Figure 4.9 – <i>Evolution of the number of employees per branch</i>	106
Figure 4.10 – <i>Managerial structure of the branch network</i>	108

Chapter 5

Figure 5.1 – <i>Decomposition of the cost-effectiveness measure</i>	124
Figure 5.2 – <i>Efficiency distributions for the production approach under CRS and VRS.</i> ..	135
Figure 5.3 - <i>Efficiency distributions for the intermediation approach (output set based on volume of business) under CRS and VRS</i>	136
Figure 5.4 - <i>Efficiency distributions for the intermediation approach (output set based on revenues) under CRS and VRS</i>	136
Figure 5.5 – <i>Intermediation efficiency versus profitability levels</i>	153
Figure 5.6 – <i>Profits levels achievable through the elimination of inefficiencies</i>	156
Figure 5.7 – <i>Operational efficiency versus profitability levels</i>	156

Chapter 6

Figure 6.1 - <i>Distribution of branches by current scale size</i>	165
Figure 6.2 - <i>Efficiency components and branches' scale size (input oriented assessment)</i> ..	167
Figure 6.3 - <i>Efficiency components and branches' scale size (output oriented assessment)</i> ..	167
Figure 6.4 - <i>Efficiency components under VRS</i>	169
Figure 6.5 – <i>Returns to scale characterisation</i>	172
Figure 6.6 – <i>Characterisation of branches' returns to scale</i>	174
Figure 6.7 - <i>Returns to scale and branches' scale size</i>	174

Chapter 7

Figure 7.1 – <i>Price scenarios considered for cost efficiency assessment</i>	189
Figure 7.2 – <i>Cost efficiency measurement</i>	191
Figure 7.3 - <i>Price (weights) space for the DEA efficiency assessment</i>	195
Figure 7.4 – <i>Optimistic CE versus Farrell CE</i>	200
Figure 7.5 – <i>Illustration of Pessimistic CE</i>	206
Figure 7.6 – <i>Production possibility set for the example considered</i>	209
Figure 7.7 – <i>Components of economic efficiency under a variable price scenario</i>	212
Figure 7.8 – <i>Difference between the Optimistic CE and Pessimistic CE measures</i>	224

Chapter 8

Figure 8.1 – <i>Illustration of the ‘Program Efficiency’ method</i>	236
Figure 8.2 - <i>Malmquist input-based productivity index</i>	241
Figure 8.3 – <i>Index for the comparison of within-region efficiency spread, using Lisbon as the reference region</i>	255
Figure 8.4 – <i>Index for the comparison of frontier productivity, with the North as the reference region</i>	257
Figure 8.5 – <i>Adjusted index for the comparison of productivity of the regional best-practice frontiers, with Lisbon as the reference region</i>	258
Figure 8.6 – <i>Adjusted index for the ranking of regional performance</i>	259

LIST OF TABLES

Chapter 2

Table 2.1 – <i>Criteria for identifying the nature of returns to scale in the ‘weights’ formulation of the DEA model</i>	38
Table 2.2 - <i>Criteria for identifying the nature of returns to scale in the ‘envelopment’ formulation of the DEA model</i>	39

Chapter 4

Table 4.1 – <i>Evolution of branch employees’ productivity</i>	106
--	-----

Chapter 5

Table 5.1 - <i>Summary information of variables related to the production approach</i>	128
Table 5.2 - <i>Summary information of variables used in the intermediation approach</i>	132
Table 5.3 – <i>Cost-effectiveness results for the intermediation approach</i>	138
Table 5.4 - <i>Results of the technical efficiency measure for the intermediation approach</i>	139
Table 5.5 - <i>Input allocative efficiency results for the intermediation approach</i>	139
Table 5.6 – <i>Output mix efficiency results for the intermediation approach</i>	140
Table 5.7 - <i>Summary of the intermediation efficiency results for the alternative output specifications</i>	141
Table 5.8 – <i>Cost efficiency results for the production approach</i>	143
Table 5.9 – <i>Pure technical efficiency for the production approach</i>	144
Table 5.10 – <i>Scale efficiency and related measures for the production approach</i>	145
Table 5.11 - <i>Input allocative efficiency for the production approach</i>	147
Table 5.12 – <i>Summary of the efficiency measures for the production and intermediation approaches</i>	148
Table 5.13 - <i>Target input and output levels for the intermediation approach</i>	155

Chapter 6

Table 6.1 - <i>Summary statistics of the input and output data</i>	163
Table 6.2 - <i>Pure technical efficiency (PTE) and scale efficiency (SE) results</i>	166
Table 6.3 – <i>Components of technical inefficiency</i>	171
Table 6.4 - <i>Targets for the branch network with restricted peers</i>	179
Table 6.5 – <i>MPSS targets for the branch network</i>	185

Chapter 7

Table 7.1 – <i>Inputs, output and prices at the DMUs</i>	190
Table 7.2 – <i>Example data set</i>	208
Table 7.3 - <i>Targets with variable prices</i>	218
Table 7.4 - <i>Targets with fixed prices</i>	218
Table 7.5 - <i>Branches' input prices in different regions of Portugal</i>	221
Table 7.6 – <i>Farrell CE measure and its components</i>	222
Table 7.7 – <i>Targets to achieve cost efficiency (with current prices)</i>	222
Table 7.8 – <i>Cost efficiency measures with price uncertainty</i>	224
Table 7.9 – <i>Efficiency measures considering variable prices</i>	225
Table 7.10 - <i>Targets to achieve economic efficiency with minimal regional prices</i>	226
Table 7.11 - <i>Targets to achieve economic efficiency with minimal changes to input mix</i> . 226	
Table 7.12 - <i>Comparison of targets at the overall network level</i>	227
Table 7.13 - <i>Comparison of targets at the branch level</i>	228

Chapter 8

Table 8.1 - <i>Summary statistics for the input/output data</i>	254
Table 8.2 – <i>Index for the comparison of within-region efficiency spread (IE_i^{AB})</i>	254
Table 8.3 - <i>Index for the comparison of productivity between two regional frontiers</i> <i>(IF_i^{AB})</i>	256
Table 8.4 - <i>Index for the ranking of frontier productivity of all regions ($IF(adj)_i^{AB}$)</i>	258
Table 8.5 – <i>Index for an overall comparison of regional performance ($I(adj)_i^{AB}$)</i>	259

Appendix 7A

Table A.1 – <i>Inputs, output and prices at the DMUs</i>	298
Table A.2 – <i>Efficiency estimates for the example considered</i>	303

ACKNOWLEDGMENTS

I would like to express my gratitude to my supervisor, Professor Robert Dyson, for his excellent ideas, advice, interest and encouragement, which were invaluable for the completion of this thesis.

I would also like to thank the DEA research group at Warwick for their useful comments made in many stimulating discussions: Rachel Allen, Laura Read, Nikos Maniadakis, Estelle Shale, Cláudia Sarrico, Ali Emrouznejad, Dr. Emmanuel Thanassoulis and Dr. Victor Podinovski.

The continuous interest and support of Prof. Rui Guimarães, from the University of Porto, is greatly appreciated.

I would also want to express my gratitude to the Portuguese Bank who supplied data and relevant information concerning the empirical part of the dissertation.

I am particularly grateful to my parents for all the motivation to pursue postgraduate studies abroad and all the support during the different phases of this research. Most of all, I would like to express my gratitude to my husband Pedro, to whom I dedicate this work, for all the support and understanding, invaluable during the process of this research.

The financial support provided by the *Ministério da Ciência e da Tecnologia, Subprograma Ciência e Tecnologia do 2º Quadro Comunitário de Apoio, Portugal*, is gratefully acknowledged.

DECLARATION

The following publication is based on the work in Chapter 6 of this thesis.

- Camanho A.S. and Dyson R.D. 1999. “Efficiency, Size, Benchmarks and Targets for Bank Branches: An application of Data Envelopment Analysis”. *Journal of the Operational Research Society*, 50 (9): 903-915.

ABSTRACT

This aim of this thesis is to develop a comprehensive methodology for assessing performance and setting targets in multi-unit organisations in the financial sector. These are structured as networks of decision making units (DMUs) that seek to operate efficiently, satisfy customer requirements effectively, and generate profit. The achievement of this objective relies on the use of the Data Envelopment Analysis (DEA) method, which is the main subject area of this thesis. It involved the development of new models and methods for performance measurement and improvement at the DMUs.

To ensure the relevance of the methodology developed, a commercial bank is used as a case study. The models and methods developed were motivated by the study of the bank branches from this institution and illustrated with empirical data. This helps to guarantee that the developments are driven from the needs of the organisations, which contributes to move the DEA method into the ‘real problem’ zone. An effort is made to ensure that the models and methods developed in this thesis are generic and applicable to other types of ‘for-profit’ organisations outside the financial services sector.

The thesis is structured as follows. An overview of frontier analysis methods, with particular emphasis on the DEA method is presented in Chapter 2. This chapter sets up the ground for the enhancements to the DEA method presented throughout the thesis. Chapter 3 reviews the literature on banking performance assessment. It summarises the main aims, methodologies and conclusions of previous research, with emphasis on the studies based on the DEA method. The information gathered is used to guide the choice of the themes and questions addressed in the context of the analysis of financial institutions’ performance. Chapter 4 introduces the commercial bank analysed and the financial sector where it operates. The description of the bank concerns the methods currently used to assess branches’ performance.

Chapter 5 develops a framework for performance appraisal, integrating efficiency and profitability dimensions. In the context of financial institutions’ assessment, it is proposed the assessment of efficiency from two different perspectives, corresponding to the operational activity and the outcomes of financial intermediation. In order to provide a comprehensive efficiency assessment, a new DEA model is used, which can identify inefficiencies in both input and output levels, considering an objective of cost minimisation. The resulting efficiency measure is decomposed in order to provide a comprehensive picture of the inefficiency sources and its managerial implications.

The following chapters explore different aspects of operational efficiency in greater detail, providing both enhanced models of efficiency measurement and target setting. Chapter 6 focuses on the analysis of the effect of scale size on efficiency. Performance improvement issues relating to the choice of appropriate benchmarks and practical aspects relating to the implementation of the DEA results are addressed. Chapter 7 focuses on the analysis of cost efficiency considering different price scenarios, including price uncertainty at the DMU level and situations where both input and price adjustments are possible.

Chapter 8 explores the differences in performance of groups of bank branches in different locations, associated with distinct environmental conditions. A new performance index is developed, which can disentangle within-group managerial inefficiencies from those attributable to the context within which the DMUs are required to operate.

Overall, this thesis contributes to illustrate the relative strengths of DEA with respect to a multitude of purposes of performance evaluation and improvement. It also provides a comprehensive assessment of a financial institution, which shows that the DEA method can be successfully used as a decision support tool for many issues faced by these organisations.

GLOSSARY

CE:	Cost Efficiency
CRS:	Constant Returns to Scale
DEA:	Data Envelopment Analysis
DFA:	Distribution Free Approach
DM:	Decision Maker
DMU:	Decision Making Unit
DRS:	Decreasing Returns to Scale
EAS:	Earning Analysis System
EU:	European Union
FDH:	Free Disposal Hull
GDP:	Gross Domestic Product
IAE:	Input Allocative Efficiency
IC-AR:	Input Cone Assurance Region
IRS:	Increasing Returns to Scale
MPSS:	Most Productive Scale Size
NDRS:	Non-Decreasing Returns to Scale
NIRS:	Non-Increasing Returns to Scale
OME:	Output Mix Efficiency
PPS:	Production Possibility Set
PTE:	Pure Technical Efficiency
SE:	Scale Efficiency
SFA:	Stochastic Frontier Approach
SIM:	System of Incentives and Motivation
TE:	Technical Efficiency
TFA:	Thick Frontier Approach
VRS:	Variable Returns to Scale

CHAPTER 1

Introduction

1.1 Introduction

The 1980s and 1990s have witnessed major structural changes in the world economy. The globalisation of markets and finance, the creation of regional economic blocks such as the European Union (EU) and the introduction of new technologies in product design and manufacturing are just some of the main features of the new international economic order.

The majority of the sectors in industrialised countries have suffered the consequences of these changes, e.g., increased competition, squeezed profitability margins, sophisticated clients requiring better quality of service, and pressures to cut prices and quickly develop and bring to market new products.

The banking industry is one of the sectors that has been hardest hit by the effects of these changes. Recent developments in financial markets, such as liberalisation, securitisation, technological evolution, globalisation, tighter competition, and the generally growing importance of financial services in the economic activity of countries have all put an increasingly sharp focus on the activities of banks. Detailed discussions on this topic can be found in Canals (1993), Dermine (1993), Molyneux *et al.* (1996) and Revell (1994).

In particular, the European financial services sector has been affected by several powerful forces of change.

First, a marked feature of banking in Europe during the 1980s and 1990s has been the liberalisation¹ and evolution towards more open markets. In particular, the creation of a single European financial market, a single European currency and the harmonisation of regulation in all EU countries have brought stronger competitive pressures to the national markets. This was specially felt in countries with a tradition of close and highly regulated financial markets, whose institutions had to adapt rapidly to keep up with the pace of changes.

Second, the financial disintermediation² and securitisation³ contributed to the growing importance of capital markets in the economy of industrialised countries. This opened the possibility of companies turning more often to capital markets for cheaper financing. Also, capital markets are attracting savers willing to invest their funds in assets that offer a better combination of risk and return. Thus, the traditional intermediation role played by banks, both in terms of deposits and lending business, has been strongly affected by the activity of capital markets.

Third, information systems and technological innovations have triggered at least two far-reaching revolutions in the financial industry. The first is the innovation in the way commercial banking, investment banking and capital market tasks are carried out. The second is that information systems have broken the space and time barriers, turning the world financial system into a truly global market. As a result, the demarcation lines between the various types of financial institutions (e.g., commercial banks, investment

¹ Liberalisation (or deregulation) is the removal of controls imposed by governments on the operation of financial markets.

² Disintermediation is the elimination of financial intermediaries, such as bankers and brokers, from transactions between borrowers and lenders or buyers and sellers in financial markets. Disintermediation has been a consequence of improved technology and liberalisation and allows both parties in a financial transaction to reduce costs by eliminating payment of commissions and fees.

³ Securitisation involves the transformation of loans into traded securities.

banks or other non-bank intermediaries such as insurance companies) have become increasingly blurred. The emergence of these new financial intermediaries, carrying out functions similar to those traditionally associated with banks, has significantly changed the structure of the financial industry.

Finally, due to the joint effect of these factors, more than ever before the banking industry is facing fierce competition.

Banking structures and strategies are now involved in a fundamental, far-reaching process of realignment and change. The financial margin of banks in all industrialised countries has fallen dramatically. The concentration has increased, as banks have sought to protect and strengthen their position through mergers and acquisitions. If past mergers were driven more by desires for size and market expansion, current mergers are driven by a greater emphasis on cost reductions, profitability and greater efficiency within the banking firm. There is an increasing pressure to use all of bank's resources to maximum advantage.

This challenging and more hostile banking environment has caused improved efficiency to be an increasingly important banking target. Under today's circumstances, the penalties for flawed evaluations of banking performance are often more immediate and may be severe.

It is important for both bankers and supervision / regulatory authorities to know whether financial institutions are becoming more efficient. At the organisational level, improved efficiency is expected to lead to greater profits, greater customer satisfaction, improved capital adequacy positions and improved risk-taking capabilities. These characteristics increase the competitiveness and economic viability of the organisation. At the industry

level, more efficient institutions lead to more flexible and robust financial systems, which are essential to the economic development of countries.

This explains why performance measurement has attracted enormous attention in recent years. However, as the role of performance measurement is rising, it is acknowledged that the traditional methods of performance appraisal currently used in most organisations have inadequacies. Some questions can be posed regarding their ability to provide relevant managerial information for today's business environment. Although traditional accounting-based methods will remain important as standard means for providing information to management, supervision authorities, shareholders and the general public, new measures of performance based on "real" resources and outcomes of the organisations are needed, founded on rigorous and sophisticated methods.

This thesis addresses the issues of performance measurement and improvement in organisations. Due to the importance of the financial sector in the world economy, allied to the substantial changes it has recently gone through, it was chosen as the underlying context for the research of this thesis.

The motivation and main objectives of this thesis are discussed in detail in the next section.

1.2 Motivation and research objectives

Although the efficiency of financial institutions has attracted considerable attention in recent years, (e.g., the recent survey by Berger and Humphrey (1997) included 130 studies that applied frontier efficiency analysis to financial institutions from 21 countries), Berger *et al.* (1993) caution that the research has not kept pace with the changes in the industry.

After the long periods of regulation and restricted competition in the financial services sector of many countries, it is likely that the institutions are not yet operating with maximum efficiency and productivity levels. However, research on the internal efficiency level of financial organisations is still limited. Most studies reported are done at the industry level, which may result in overestimation of the institutions' efficiency. This is because the best-practice institutions identified in these conventional studies almost certainly contain internal inefficiencies that are not accounted for. It is important to estimate efficiency within the organisations to complement the industry-level assessments.

The increased competition in financial markets motivated the institutions to strive for efficiency in order to cut costs and deliver better customer services. However, the performance appraisal methods currently available at the organisations are not enough for providing the information required for efficiency improvement. More sophisticated and robust methods for benchmarking and dissemination of best practices must be developed and made available to the organisations.

The broad subject area of this thesis is the use of frontier analysis methods, in particular the **Data Envelopment Analysis (DEA)** method, for the assessment and improvement of performance in organisations. The main advantages of the use of frontier methods are that they can explicitly consider the use of multiple inputs to produce multiple outputs, the construction of the best-practice frontier is based on empirical data (measured in physical or monetary units), and they rely on powerful and comprehensive optimising techniques. From the alternative frontier methods available, Data Envelopment Analysis was the chosen method for this thesis. Given the context of financial institutions' assessment, the DEA method was considered the most appropriate due to the greater

flexibility of the assumptions imposed on the frontier estimated and the possibility to decompose the efficiency estimates into the different sources in a simple manner. The motivation for using the DEA method is described in greater detail in Chapter 3.

The main objective of this thesis is to provide a comprehensive framework for performance measurement and improvement in financial organisations. To ensure the relevance of the models and methods developed within this framework, a branch network from a commercial bank operating in Portugal is analysed. This guarantees that the developments are driven from the real needs of today's organisations, i.e., *application driven theory*.

The following research questions related to the performance of bank branch networks will be addressed in this thesis:

- Can an overall framework for the assessment of performance be developed, integrating efficiency and profitability dimensions (Chapter 5)?
- Is it possible to characterise the impact of scale size on efficiency and specify targets that eliminate managerial and scale inefficiencies without changing significantly the profile of the productive units (Chapter 6)?
- Is it possible to determine the extent to which a productive unit can use resource levels of lower aggregate cost when exact input prices are not known and only ranges of acceptable prices can be defined? Can the optimal balance of resources for production at minimal cost be defined in complex scenarios where both the input and price levels of a productive unit can be adjusted (Chapter 7)?
- Can the performance of groups of productive units operating under different conditions be compared, separating managerial inefficiencies from those attributable to the environment within which the units operate (Chapter 8)?

Throughout the thesis, an effort is made to ensure that the methods developed within the context of performance measurement and improvement in bank branch networks are generic and applicable to other types of “for-profit” organisations outside the financial services sector. New methods for benchmarking best practice and improving

performance are important to all organisations. They can contribute to the strengthening of the competitive position of the organisations, which is essential to guarantee viability in the modern day world.

1.3 Thesis structure

The structure of this thesis is as follows. The next chapter outlines the state of the art in terms of frontier analysis methods for the evaluation of performance. Particular emphasis is given to the DEA method, which was considered the most appropriate method for the achievement of the research objectives stated for this thesis.

Chapter 3 reviews the literature on banking efficiency measurement, focusing on the use of the DEA method.

Chapter 4 provides the background to the Portuguese financial services sector and introduces the bank analysed. Particular emphasis is given to the description of the performance measurement methods currently used in this bank.

Chapter 5 develops a general framework for performance appraisal that integrates efficiency and profitability measures. A comprehensive efficiency assessment of the bank branch network is described, including the analysis of the inefficiency sources. Branches' efficiency is assessed under two perspectives, corresponding to the operational and intermediation activities.

Chapters 6 and 7 explore in greater detail branches' operational activity. Chapter 6 focuses on the analysis of the effect of scale size on efficiency. Performance improvement issues relating to the choice of appropriate benchmarks and practical issues relating to the implementation of the DEA results are addressed. Chapter 7 focuses on the analysis of the economic aspects of efficiency, considering assessments with

different price scenarios. It explores scenarios of price uncertainty at the DMU level and situations where both input and price adjustments are possible.

Chapter 8 compares the performance of branches located in different regions, with distinct environmental characteristics. The resulting information contributes to the clarification of the causes for different performance levels among the branches, i.e., inadequacies in managerial approaches or less favourable environmental conditions that affect branches' productivity.

Finally, Chapter 9 presents the conclusions and proposes future research directions. It summarises the main contributions of the thesis and discusses the extent to which the thesis objectives were achieved.

CHAPTER 2

The assessment of performance

2.1 Introduction

The purpose of this chapter is to introduce the main concepts and methods for the evaluation of performance in organisations. The related literature is vast and rooted both in economics and management science. Although an effort is made to provide a comprehensive introduction to performance measurement, the presentation is selective and focuses on **frontier analysis** methods. Particular emphasis is given to the **Data Envelopment Analysis** (DEA) method and the concepts of **relative efficiency**, which are crucial for the achievement of the research objectives stated for the thesis.

The efficiency of a productive unit, referred to as a **Decision Making Unit** (DMU), is defined by comparing its **inputs** and **outputs** to those of the best performing from its peers. The inputs correspond to the resources used, whereas the outputs are the products or services obtained as a result of the production process. The level of outputs produced must be related in some way to the level of inputs used to secure them. This relation is called the **technology of production** and defines the maximum possible output obtainable from given inputs.

Exact knowledge of the technology of production is not usually available. Thus, for a long time economists and management scientists have developed alternative methods for deriving empirically the technology of production from a set of similar DMUs under analysis. Despite the differences in the methods available for the estimation of the

technology of production, efficiency is always defined by comparing observed to optimal productive performance.

This chapter starts with a brief historical overview on the measurement of efficiency, focusing on the evolution of frontier analysis methods. As implied by their name, frontier methods estimate production technologies that go through the boundary of the production space. For this reason, they are deemed the most appropriate for the assessment of efficiency, as they are based on ‘best practices’ rather than ‘average performance’.

This chapter provides an overview of the main frontier methods, including those most frequently used in the analysis of performance of financial institutions. The underlying characteristics of these methods are described in order to highlight their relative strengths and weaknesses.

The core of this chapter provides an introduction to the DEA method. It includes a description of the theory underlying the representation of the technology of production and the efficiency frontier in DEA, which is based on the Axiomatic Approach (Koopmans, 1957; Debreu, 1951; Shephard, 1970). The main DEA models for the evaluation of efficiency are reviewed, before discussing the recent development in the DEA literature.

More detailed discussions on the material presented in this chapter can be found in Banker *et al.* (1989), Bauer (1990), Boussofiane *et al.* (1991), Charnes *et al.* (1994), Charnes and Cooper (1985), Cooper *et al.* (1996), Fare, Grosskopf and Lovell (1994), Fried *et al.* (1993), Seiford and Thrall (1990).

This chapter is structured as follows. Section 2.2 describes the historical evolution of frontier analysis. Section 2.3 provides a brief overview of the main frontier analysis methods. As DEA is the broad subject area of this thesis, section 2.4 provides a more detailed introduction to the DEA method and discusses the extensions that are most relevant to this thesis. Section 2.5 summarises and concludes.

2.2 Historical evolution of frontier analysis

Traditional approaches to efficiency measurement consist of a comparison between observed and optimal values of the outputs, or the inputs, of a decision making unit (DMU). The comparison can take the form of observed to maximum output obtainable from the given input, or the ratio of minimum input required for producing the given output to the observed input. In these two comparisons the optimal is defined in terms of the physical production possibilities, and efficiency is called technical.

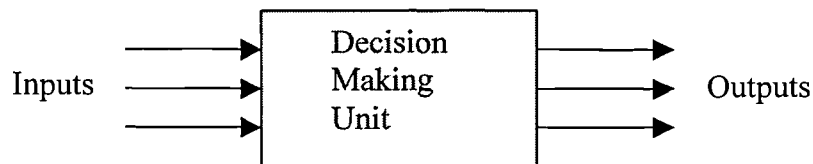


Figure 2.1 - *The production process*

It would also be possible to define the optimal incorporating the economic goal of the DMU. In this case, efficiency is called economic and is measured by comparing observed and optimum cost, revenue or profit, subject to appropriate constraints both on quantities (i.e., reflecting the technology of production) and prices (i.e., reflecting the market conditions).

Even at this conceptual stage of efficiency measurement two problems arise: How many and which inputs and outputs should be included in the analysis, and how should the *optimal* production possibility of the DMU be determined?

In relation to the first problem, it is clear that the efficiency results obtained are highly dependent on the selection of variables to be included in the assessment, as well as how they are measured. These variables should be chosen to reflect the primary aims of the assessment. For example, when assessing the performance of schools, one can examine the ability of individual schools to utilise their resources in order to achieve high examination results. In this case, it would be appropriate to choose as inputs the resources available at a school (e.g., number of teachers, facilities and expenditure) and as outputs the examination achievements of pupils. Conversely, if the objective of the assessment concerned the value added at schools, the inputs should include information on the entry standards and socio-economic background of pupils (see Thanassoulis and Dunstan, 1994). The issue of variable selection and measurement will be explored in greater detail at the end of this chapter, within the context of the DEA method.

The second problem, relating to the determination of the *optimal* production capacity of a DMU, is the most difficult to answer. Traditional economic approaches theoretically define a production function, which is a mathematical representation of the relation between inputs and outputs, and is defined as the maximal possible output obtainable from given inputs. The seminal work by Cobb and Douglas (1928), relating to the estimation of average production functions, contributed substantially to the development of this field of economics. Since then, more flexible production function forms were developed and tested on empirical data. However, although the estimation of average production functions has become commonplace in economics, the estimation of frontier production functions has only attracted widespread attention recently. As Aigner *et al.* (1977, p21) mentioned:

“ The theoretical definition of a production function expressing the maximum amount of output obtainable from given input bundles with fixed technology has been accepted for many decades. And for almost as long, econometricians have been estimating average production functions.”

Despite the key contributions from economic theory to frontier analysis, for many years the productivity literature ignored the efficiency component due to the difficulties in estimating optimal, as opposed to average, input-output relations.

The underpinnings of efficiency measurement date back to the work of Debreu (1951) and Koopmans (1957). Debreu provided the first measure of efficiency, which was called the ‘*coefficient of resource utilisation*’ and Koopmans was the first to define the concept of technical efficiency. Farrell (1957) extended their work in a seminal paper whose key development was to show how to bring data to bear on Debreu’s formulation of ‘*coefficient of resource utilisation*’. This changed the focus from absolute to relative efficiency.

Farrell (1957) work provided the foundations to the estimation of empirical frontier production functions. In most production processes, the conversion of inputs into outputs does not follow a known functional form. Therefore, the traditional economic method, based on theoretically defined production functions requiring a-priori specification of a functional form, is likely to identify as best performance some unattainable ideal. Farrell (1957) suggested changing the focus from absolute to relative efficiency by promoting the comparison of a DMU to the best actually achieved by peers performing a similar function.

Farrell (1957) graphical illustration of the efficiency concepts has now become classical. In order to provide a pictorial representation of his ideas, consider a set of DMUs that produce a single output (Y) using two inputs (X_1 and X_2) in varying quantities, as shown in Figure 2.2.

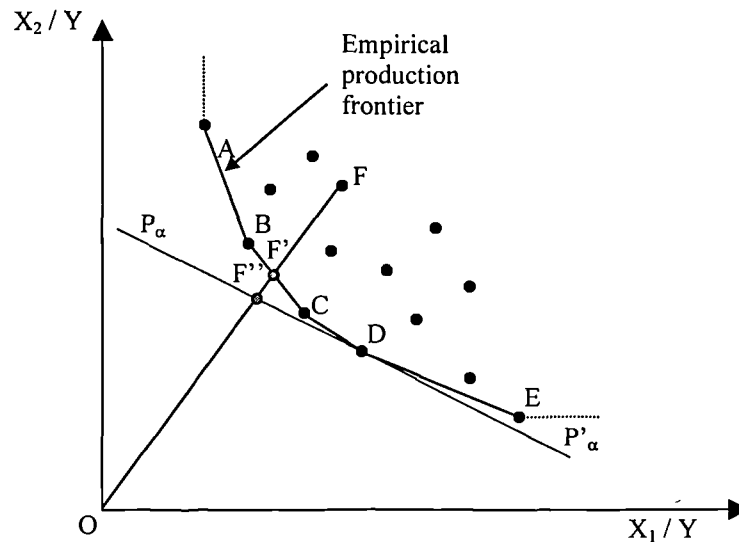


Figure 2.2 - Efficiency measurement

Farrell (1957) analysis had an input-reducing focus and assumed constant returns to scale¹ (CRS). As the DMUs are each producing a normalised level of output, this allows their representation in a two-dimensional diagram. The segments linking DMUs A, B, C, D and E form the technical efficient production frontier. Note that the frontier has a piecewise linear shape.

DMU F will be used to illustrate the efficiency concepts. Its technical efficiency² is given by the ratio $\frac{OF'}{OF}$. A ratio less than one indicates that it is possible to build a composite DMU that employs the same proportion of inputs (or input mix) and can produce the same output as the assessed DMU using only a fraction of its inputs.

Looking beyond technical efficiency, Farrell (1957) also proposed a measure of economic efficiency based on a cost minimising behaviour. The measure of Farrell's cost efficiency is illustrated for DMU F. It requires the specification of an isocost³ line,

¹ Under efficient input to output transformations, CRS means that scaling the input levels by a factor α leads to an equally proportionate scaling in the outputs by the same factor α .

² Note that Farrell (1957) measure of technical efficiency is the inverse of the distance function, introduced by Shephard (1970).

³ An isocost is a line in which all points have the same cost value.

whose slope is equal to the observed prices ratio at the DMU (i.e., $-P_1/P_2$, where P_1 is the price of input 1 and P_2 the price of input 2). This is represented in Figure 2.2 by the line $P_\alpha P'_\alpha$. Comparing points F' and D on the production frontier, although they both exhibit 100% technical efficiency, the costs of production at D will only be a fraction $\frac{OF''}{OF'}$ of those at F' . This ratio is defined as the input allocative efficiency of DMU F .

Input allocative efficiency attempts to capture the inefficiency arising solely from the wrong choice of technically efficient input combinations given input prices, i.e., measures the extent to which a DMU uses the various factors of production in the best proportions in the light of their prices. If DMU F were perfectly efficient, both technically and allocatively, its costs would be a fraction $\frac{OF''}{OF}$ of their current level.

This ratio gives a measure of cost efficiency. It indicates the extent to which the DMU is supporting its current level of outputs at minimum cost.

In summary, the work of Farrell (1957) was innovative for a number of reasons. It relaxed the need for specifying a functional form prior to estimating efficiency from empirical data. It introduced the principle of constructing a hypothetical DMU (such as F') as a convex combination of observed DMUs. It recognised the existence of multi-input and multi-output production technologies without, however, providing a method for the estimation of the production frontier⁴.

⁴ Farrell (1957) did not suggest the use of linear programming in the 1957 paper.

Despite Farrell (1957) developments, efficiency and its measurement only attracted attention again much later. After Farrell (1957), the evolution in the assessment of productive efficiency came via two parallel routes that differ in the way the frontier is specified and estimated. The frontier can be specified as parametric or non-parametric. Both methods can be further divided into stochastic and deterministic. To estimate the frontier, statistical or mathematical programming techniques can be used. Figure 2.3 shows the various types of production frontiers.

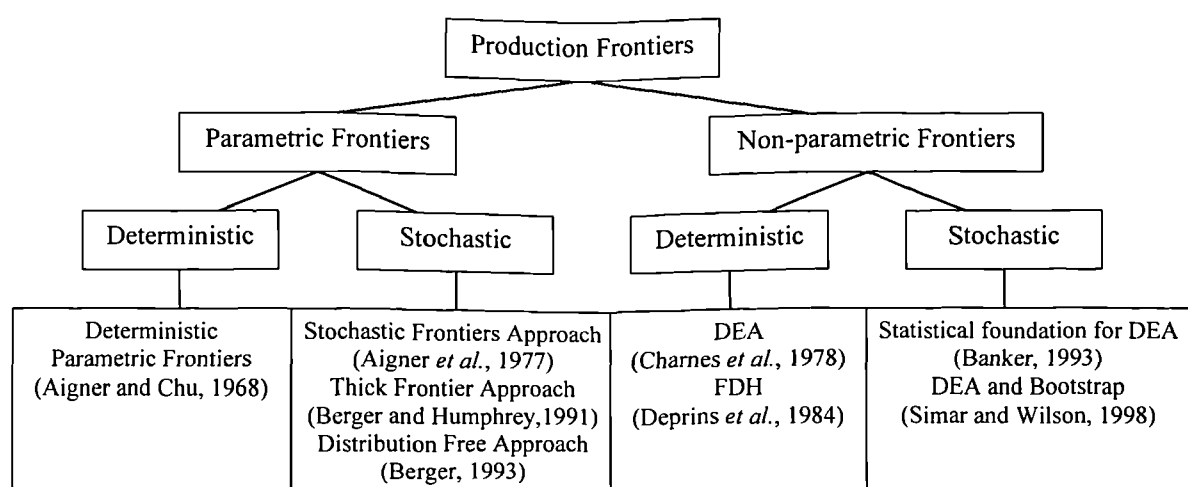


Figure 2.3 - Classification of production frontiers

Under the parametric approach, the technology (e.g., represented by a production or cost function) is specified as a function with a precise mathematical form (e.g., the translog or the Cobb-Douglas function). The type of function representing the frontier must be specified *a-priori*, and its parameters are estimated from the empirical data.

Under the non-parametric approach, the technology is defined by a set of properties that the points in the production possibility set (PPS) are assumed to satisfy. No function with constant parameters needs to be specified⁵.

⁵ The specification of production sets is based on the Axiomatic approach (Koopmans, 1957, Debreu, 1959, Shephard, 1970)

A deterministic approach assumes that all the deviations of observed production from the estimated frontier are exclusively explained by inefficiency. It is assumed that there are no random factors affecting the construction of the frontier, such as random noise or errors in the data. Thus, all observations must lie on or below the frontier. The estimation of deterministic frontiers involves the use of mathematical programming techniques.

The stochastic approach allows for random noise and measurement error in the data. These factors may affect the DMUs performance and be responsible, together with inefficiency, for observed deviations from the frontier. As a result, the DMUs may lie above or below the frontier, due to either inefficiency or random error. The stochastic approach involves the use of statistical techniques.

The next section provides a brief overview of the main frontier methods. From the various methods that have been developed, the deterministic non-parametric approach has seen the most development and a substantial body of work has used it. As DEA is the broad subject area of this thesis, the final section of this chapter is dedicated to its description.

2.3 Overview of frontier analysis methods

2.3.1 Parametric deterministic frontiers

The deterministic parametric frontier method (Aigner and Chu, 1968) requires the prior specification of a functional form for the technology of production. The parameters of the function are calculated from empirical data with the application of mathematical programming techniques. Using this approach, the estimation of technical efficiency for a single-output multiple-input situation requires the definition of a production function.

Similarly, the estimation of cost efficiency for a single input (measured by total cost) multiple-output situation requires the specification of a cost function⁶. Conversely, in the case of multiple-input and multiple-output situations, the efficiency measure has to be obtained from the estimation of a distance function⁷, (see Fare *et al.* (1993) for an application of this approach).

2.3.2 Parametric stochastic frontiers

There are three main parametric stochastic frontier approaches. They all specify a functional form for the frontier and differ on how the stochastic component is modelled.

The Stochastic Frontier Approach (SFA) developed by Aigner, Lovell and Schmidt (1977) specifies a functional form for the cost, profit⁸ or production function, and allows for random noise such as luck or specification error. The SFA assumes that deviations from the estimated frontier are composed by inefficiencies and random error. The inefficiencies are assumed to follow a one-sided distribution, usually the half-normal, while random errors follow a symmetric distribution, usually the standard normal. The inefficiencies must have a truncated distribution because they cannot be negative. For further details on the estimation of stochastic frontiers see Forsund *et al.* (1980), Schmidt (1985), Lovell and Schmidt (1988) and Bauer (1990).

⁶ A cost function gives the minimal cost level at which it is possible to produce the outputs, given input prices.

⁷ The (input) distance function gives the maximal factor by which a given input vector can be reduced radially within the production possibility set. See Shephard (1953, 1970) for a mathematical definition of the distance function.

⁸ The profit function gives the maximum profit that can be obtained given input and output prices.

The Distribution-Free Approach (DFA) developed by Berger (1993) also specifies a functional form for the frontier but makes no strong assumptions regarding the specific distributions of the inefficiencies or random errors. This approach is only applicable when panel data⁹ is available. The DFA assumes that the efficiency of each DMU is stable over time, whereas random error tends to average out to zero over time. The estimate of inefficiency for each DMU in a panel data set is then determined as the difference between its average residual and the average residual of the DMUs on the frontier, with some truncation performed to account for the failure of the random error to average out to zero fully. Using this approach, the inefficiencies can follow any distribution, even one that is close to symmetric, as long as the inefficiencies are non-negative. Although with this approach specific distributional assumptions may be avoided, it is still imposed that efficiency is stable over time. This is a very strong assumption, particularly as the time period under analysis increases.

The Thick Frontier Approach (TFA) developed by Berger and Humphrey (1991) also specifies a functional form for the frontier. Instead of estimating a precise frontier edge, this method estimates a cost function for the lowest average cost quartile of DMUs, which may be thought of as a 'thick frontier'. It is assumed that the DMUs in this lowest average cost quartile are of greater than average efficiency, whilst the error term obtained from the estimation of this cost function is assumed to reflect only measurement error and luck. This method also involves the estimation of a cost function for the DMUs in the highest average cost quartile, in which it is assumed that the DMUs are of less than average efficiency. The sample is generally divided into size classes prior to forming the cost quartiles to ensure a reasonable representation of all sizes of DMUs across quartiles.

⁹ Panel data refers to the pooling of observations on the same set of DMUs over several time periods.

The differences in predicted cost obtained from the two cost functions defined, evaluated at the mean of the size classes defined, are separated into ‘market factors’, explained by the differences in the available exogenous variables, and an ‘inefficiency residual’. The ‘inefficiency residual’ can then be decomposed into several types of inefficiencies.

The advantages of the TFA are that it does not impose any distributional assumptions on either inefficiency or random error and reduces the effect of extreme observations in the data. The main assumptions of this method are that the error term within the lowest and highest cost quartiles reflects only random measurement error or luck, while the differences between the lowest and highest cost quartiles reflect only inefficiencies and the effect of market factors. Its main limitation is that it does not provide point estimates of inefficiency for individual DMUs. It is intended to provide an estimate of the general level of inefficiency for all DMUs.

2.3.3 Non-parametric deterministic frontiers

The most widely used non-parametric techniques are Data Envelopment Analysis (DEA) developed by Charnes, Cooper and Rhodes (1978), and Free Disposal Hull (FDH), developed by Deprins, Simar and Tulkens (1984).

Data Envelopment Analysis (DEA) does not require the explicit specification of the form of the underlying production technology, as the frontier is formed by the piecewise linear segments that connect the set of frontier observations. In DEA the frontier observations are those for which no other DMU or linear combination of DMUs has more of at least one output and as much of all other outputs (given inputs) or less of at least one input and no more of all other inputs (given outputs). The efficiency measure is obtained with the application of mathematical programming techniques. Beyond the advantages of not requiring *a-priori* specification of the functional form for the frontier and enabling an

easy computation of efficiency scores using linear programming, the DEA method enables a decomposition of the efficiency measure into several components. This provides an aid to management in its search for inefficiency sources. The DEA method is described in greater detail in section 2.4.

The Free Disposable Hull (FDH) is a special case of the DEA model, where the convex combinations of the frontier observations are not part of the frontier. Thus, the difference between DEA and FDH results from the removal of the convexity assumption in the latter. From the perspective of input requirements to produce a given output, DEA assumes that linear substitution is possible between observed input combinations. In contrast, FDH assumes that no substitution is possible so that the frontier looks like a step function formed by the intersection of lines parallel to the axes drawn from the frontier observations. The FDH production possibility set is composed only of the DEA vertices and the free disposable hull points interior to those vertices. For a detailed description of this technique see Tulkens and VandenEeckaut (1995) and DeBorger *et al.* (1998).

2.3.4 Non-parametric stochastic frontiers

A key drawback of the non-parametric deterministic approaches is that they assume that there is no random noise or measurement error in the data. In order to overcome this limitation, two research routes are being pursued on the development of a statistical theory for DEA.

One is analytical and seeks to provide a statistical foundation for DEA (see Banker, 1993 and 1996). This work provides a theoretical foundation for statistical hypothesis testing in DEA. The analytical research has demonstrated that if the number of DMUs assessed

is large and given certain plausible assumptions¹⁰ concerning the structure of the technology and the distribution of the ‘true’ inefficiencies, the empirical estimates obtained from a DEA model have the following properties:

- The DEA estimator of inefficiency is a statistically consistent estimator for the true inefficiency of a DMU;
- The DEA estimator is a maximum likelihood estimator;
- The empirical distribution of the DEA inefficiency estimate for individual DMUs recovers the true inefficiency distribution for the set analysed.

The other research agenda is empirical (see Simar and Wilson, 1998), and seeks to develop and implement a stochastic version of DEA. A resampling technique, such as bootstrapping¹¹, is one way to obtain empirically the true distribution underlying the DEA efficiency estimates. Once the underlying distribution is approximated, statistical inference can be conducted.

2.4 Data Envelopment Analysis

2.4.1 Production possibility set and efficient frontiers

The theory of production underlying the DEA methodology draws on the axiomatic approach (Koopmans, 1957; Debreu, 1951, Shephard, 1970), which is based on production sets. The aim of this section is to characterise the production possibility set underlying the DEA method and, in particular, to define its efficient subset.

¹⁰The assumptions are embodied in five postulates, defined in Banker (1996, p.141), relating to the structure of the production possibility set and the probability density function for the inefficiency (i.e., *Monotonicity, Convexity, Envelopment, Likelihood of efficient performance* and *Decreasing probability density*).

¹¹The bootstrap method is a well-established computationally intensive statistical resampling method used to perform inference in complex problems. See Efron and Tibshirani (1993) for a presentation of the method.

Consider a set of $j = 1, \dots, n$ DMUs and each uses inputs $X \in \mathbb{R}_+^m$ to produce outputs $Y \in \mathbb{R}_+^s$. Thus, DMU j uses amount x_{ij} of input i , $i = 1, \dots, m$, to produce amount y_{rj} of output r , $r = 1, \dots, s$.

The production possibility set Φ contains all input-output feasible combinations corresponding to a certain production process. Formally this can be stated as follows:

$$\Phi = \{(X, Y) \mid \text{Input vector } X \text{ can produce the output vector } Y\}. \quad (2.1)$$

The following properties are postulated for the production possibility set (see Banker *et al.*, 1984; Banker and Thrall, 1992):

Postulate 1 (Inclusion of observations)

The observed $(X_j, Y_j) \in \Phi$, for all $j = 1, \dots, n$.

Postulate 2 (Inefficiency)

(a) If $(X, Y) \in \Phi$ and $X' \geq X$, then $(X', Y) \in \Phi$.

(b) If $(X, Y) \in \Phi$ and $Y' \leq Y$, then $(X, Y') \in \Phi$.

Postulate 3 (Ray Unboundedness)

If $(X, Y) \in \Phi$ then $(kX, kY) \in \Phi$, $\forall k > 0$.

Postulate 4 (Convexity)

If $(X_j, Y_j) \in \Phi$, $j = 1, \dots, n$, and $\lambda_j \geq 0$ are nonnegative scalars such that

$$\sum_{j=1}^n \lambda_j = 1, \text{ then } \left(\sum_{j=1}^n \lambda_j X_j, \sum_{j=1}^n \lambda_j Y_j \right) \in \Phi.$$

Postulate 5 (Minimum extrapolation)

Φ is the intersection of all $\hat{\Phi}$ satisfying Postulates 1, 2, 3 and 4.

In many cases, it is important to know how the technology behaves in changes of the scale of operation. This notion is captured by the returns to scale admitted by the technology of production. Returns to scale are a characteristic of the boundary of the technology of production. For input-output vectors inside the boundary, returns to scale are measured with reference to a corresponding boundary point. Constant Returns to Scale (CRS) occurs if output increases proportionally to input, Non-Decreasing Returns

to Scale (NDRS) occurs if output increases proportionally more than input and Non-Increasing Returns to Scale (NIRS) occurs if output increases proportionally less than input. These concepts are mathematically defined below (see Seiford and Zhu, 1999):

- Constant Returns to Scale (CRS) occurs if:
for $\mu > 0 \Rightarrow \mu\Phi = \Phi$.
- Non-Increasing Returns to Scale (NIRS) occurs if:
for $0 < \mu \leq 1 \Rightarrow \mu\Phi \subseteq \Phi$ or, equivalently
for $\mu \geq 1 \Rightarrow \mu\Phi \supseteq \Phi$.
- Non-Decreasing Returns to Scale (NDRS) occurs if:
for $0 < \mu \leq 1 \Rightarrow \mu\Phi \supseteq \Phi$ or, equivalently
for $\mu \geq 1 \Rightarrow \mu\Phi \subseteq \Phi$.
- Variable Returns to Scale (VRS) occurs if:
 Φ exhibits NIRS, NDRS or any combination of the above.

Postulates 1, 2, 3, 4 and 5 can be used to define a Constant Returns to Scale (CRS) production possibility set, as shown below in (2.2).

$$\Phi_{\text{CRS}} \equiv \left\{ (X, Y) \in \mathcal{R}^+ \mid X \geq \sum_{j=1}^n \lambda_j X_j, Y \leq \sum_{j=1}^n \lambda_j Y_j, \lambda_j \geq 0 \right\} \quad (2.2)$$

Exclusion of postulate 3 will lead to the definition of a Variable Returns to Scale (VRS) production possibility set, as shown below in (2.3).

$$\Phi_{\text{VRS}} \equiv \left\{ (X, Y) \in \mathcal{R}^+ \mid X \geq \sum_{j=1}^n \lambda_j X_j, Y \leq \sum_{j=1}^n \lambda_j Y_j, \sum_{j=1}^n \lambda_j = 1, \lambda_j \geq 0 \right\} \quad (2.3)$$

The CRS and VRS production possibility sets in (2.2) and (2.3) have in common a fundamental feature that includes as members of the PPS linear combinations of inputs and outputs of observed DMUs (see *convexity* postulate). Depending on whether the *ray unboundedness* postulate is included, the shape of the frontier of the PPS is affected. In the case of CRS, the frontier is defined as a conical hull, whilst in the case of VRS, due

to the exclusion of the *ray unboundedness* postulate, the frontier is defined as a convex hull of the production possibility set.

Each production possibility set has a frontier, which is defined from a subset of DMUs that satisfy the property of efficiency. In an input saving sense, Farrell (1957) notion of technical efficiency ($E_{j_0}^F$) for a DMU j_0 under CRS can be defined as follows:

$$E_{j_0}^F = \min \left\{ \theta \mid X \leq \theta X_{j_0}, Y \geq Y_{j_0}, \text{ and } (X, Y) \in \Phi_{\text{CRS}} \right\}. \quad (2.4)$$

All production possibilities (X_j, Y_j) with $\theta = 1$ are called Farrell-efficient and belong to the frontier of the PPS. Therefore, according to Farrell (1957) criteria, a DMU is technically efficient, in input terms, if it is not possible to reduce its inputs proportionally without decreasing at least one output. Alternatively, stated in output terms, a DMU is technically efficient if it is not possible to increase the outputs proportionally without increasing at least one input.

However, Farrell (1957) definition of efficiency is not sufficient for defining ‘truly’ efficient DMUs. Koopmans (1957, p.60) defined technical efficiency as follows:

“A producer is technically efficient if an increase in any output requires a reduction in at least one other output or an increase in at least one input, or if a reduction in any input requires an increase in at least one other input or a decrease in at least one output.”

The mathematical expression of this definition is as follows:

$$E_{j_0}^K = \max \left\{ s + d \mid X \leq X_{j_0} - s, Y \geq Y_{j_0} + d \right. \\ \left. \text{and } (X, Y) \in \Phi_{\text{CRS}}, s \in \mathcal{R}_+^m, d \in \mathcal{R}_+^s \right\}. \quad (2.5)$$

A value of $E_{j_0}^K = 0$ indicates that the assessed DMU j_0 is efficient in Koopmans’ sense.

The DMUs that satisfy this criteria constitute the efficient subset from the frontier of the production possibility set.

The distinction between Farrell (1957) and Koopmans (1957) notions of efficiency is illustrated pictorially in Figure 2.4.

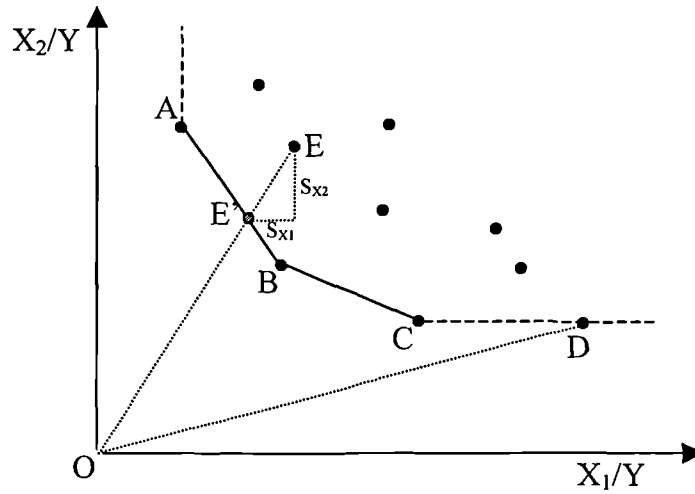


Figure 2.4 – Farrell's versus Koopmans' notions of efficiency

Farrell's efficiency is based on the radial contraction factor θ (see formula (2.4)). This implies that after the equiproportional reduction of the inputs of DMU j_0 by the factor θ , at the boundary, for at least one input, there is no scope for further reduction. Koopmans' efficiency investigates further the potential reduction of each input beyond the radial contraction factor θ .

In Figure 2.4 the frontier of the production possibility set is defined by the segments linking DMUs A, B and C and the extensions parallel to the axes spanning from A and C. In the case of DMU E, both Farrell's and Koopmans' efficiency criteria classify it as an inefficient DMU. Based on Farrell's criteria $\theta_E = \frac{OE'}{OE} < 1$ (see formula (2.4)), and based on Koopmans' criteria $s_{x_1} + s_{x_2} > 0$ (see formula (2.5)). Similarly, the DMUs on the frontier of the PPS consisting of the segments between A, B and C would be identified as efficient under both criteria. However, the two criteria differ for any DMU on the expansion of the frontier of the PPS parallel to the axes. For example, in the case

of DMU D the Farrell test will give an efficiency value equal to one, as the OD ray from the origin meets D without any interference from the efficient frontier. Thus, DMU D is ‘Farrell-efficient’. But should DMU D be considered efficient? Clearly not, as DMU C uses the same amount of input 2 and less amount of input 1 to produce the same amount of output. Using Koopmans’ efficiency criteria, it is clear that DMU D is inefficient, as it is possible to reduce the usage of input 1 from the level at D to the level at C. Farrell (1957) noticed the problem caused by DMUs like D, which he called ‘DMUs at infinity’, without however providing any methods for identifying their true efficiency.

In the remainder of this thesis, the notion of efficiency adopted will always correspond to those DMUs that satisfy the Koopmans’ criteria.

2.4.2 The approach

DEA is a linear programming technique for measuring the relative efficiency of a homogeneous set of DMUs in their use of multiple inputs to produce multiple outputs. DEA identifies a subset of efficient ‘best practice’ DMUs and, for the remaining DMUs, their efficiency level is derived by comparison to a frontier constructed from the ‘best practice’ DMUs.

DEA derives a single summary measure of efficiency for each DMU. This measure can be obtained from two perspectives, corresponding to an input-reduction or output-expansion orientation, as follows:

- Input orientation: Is the DMU using the minimum amount of the inputs given the output levels it is currently producing?
- Output orientation: Is the DMU producing the maximum amount of the outputs from its current input levels?

The choice of orientation will depend on the context of the assessment and the aims of the organisation. The efficiency measure derived with an input orientation corresponds to

the minimal factor by which all inputs of the DMU under assessment can be decreased proportionally without decreasing the level of any outputs. Conversely, the efficiency measure with an output orientation is the inverse of the maximum factor by which all outputs can be raised equiproportionally without increasing the level of any inputs. It should be noted that under CRS the measures of input and output efficiency are equivalent (see Charnes *et al.*, 1978).

Beyond the efficiency measure, DEA also provides other sources of managerial information relating to the DMUs' performance. DEA identifies the efficient peers for each inefficient DMU. This is the set of relative efficient DMUs to which an inefficient DMU has been directly compared in the derivation of its efficiency score. Therefore, DEA can be viewed as a benchmarking technique, as it allows decision makers to locate and understand the nature of the inefficiencies of a DMU by comparing it with a selected set of efficient DMUs with a similar profile.

DEA also provides information about the targets that would render an inefficient DMU efficient. These targets correspond to the input reductions and output expansions required for producing on the efficient frontier.

2.4.3 The DEA models

Having defined the production possibility set and the efficient frontier, and introduced the main features of the DEA approach, the next step is to describe the DEA model. The DEA model was first developed by Charnes, Cooper and Rhodes (1978). Its mathematical representation will be introduced via the most intuitive formulation, corresponding to a ratio model.

2.4.3.1 The ratio model

Consider a set of n DMUs, $j = 1, \dots, n$, each consuming m inputs, x_{ij} , $i = 1, \dots, m$, to produce s outputs, y_{rj} , $r = 1, \dots, s$. For each DMU j_0 under assessment, it is possible to obtain a measure of relative efficiency defined by the ratio of all outputs (y_{rj_0}) to all inputs (x_{ij_0}). The multiple inputs and outputs are reduced to a single input value and a single output value by the allocation of weights to each input and output. These weights are not defined *a-priori*, and they are chosen in order to show the efficiency of DMU j_0 in the best possible light.

For an input oriented assessment under CRS, the relative efficiency of a DMU j_0 is obtained from the following model:

DEA ratio model	(2.6)
$\text{Max } \frac{\sum_{r=1}^s \mu_r y_{rj_0}}{\sum_{i=1}^m v_i x_{ij_0}}$ $\text{s.t. } \frac{\sum_{r=1}^s \mu_r y_{rj}}{\sum_{i=1}^m v_i x_{ij}} \leq 1, \quad j = 1, \dots, n$ $\mu_r \geq \varepsilon, \quad r = 1, \dots, s$ $v_i \geq \varepsilon, \quad i = 1, \dots, m$	

μ_r and v_i stand for the output and input weights, respectively. In order to show the efficiency of DMU j_0 in the ‘best possible light’, the ratio of weighted outputs to weighted inputs is maximised, subject to the constraints that all efficiency measures for

the other DMUs must be less than or equal to unity when evaluated with similar weights.

ϵ is a mathematical infinitesimal, which ensures that the weights are strictly positive.

Model (2.6) is a fractional model but can be converted into linear form through a simple transformation, see Charnes *et al.* (1978). The next sections describe the linear programming models for computing efficiency within the DEA framework.

2.4.3.2 The DEA model with constant returns to scale

Assuming constant returns to scale, the efficiency of DMU j_0 can be determined either under input reduction or output expansion orientations, corresponding to formulations (2.7) and (2.8). These models result from the linearisation of a fractional model such as (2.6).

DEA input oriented 'weights' model under CRS		(2.7)
$\text{Max } e_{j_0} = \sum_{r=1}^s u_r y_{rj_0}$		
$\text{s.t. } \sum_{i=1}^m v_i x_{ij_0} = 1$		
$\sum_{r=1}^s u_r y_{rj} - \sum_{i=1}^m v_i x_{ij} \leq 0, \quad j = 1, \dots, n$		
$u_r \geq \epsilon, \quad r = 1, \dots, s$		
$v_i \geq \epsilon, \quad i = 1, \dots, m$		

DEA output orientated ‘weights’ model under CRS	(2.8)
$\text{Min } h_{j_0} = \sum_{i=1}^m v_i x_{ij_0}$ $\text{s.t. } \sum_{r=1}^s u_r y_{rj_0} = 1$ $\sum_{r=1}^s u_r y_{rj} - \sum_{i=1}^m v_i x_{ij} \leq 0, \quad j = 1, \dots, n$ $u_r \geq \varepsilon, \quad r = 1, \dots, s$ $v_i \geq \varepsilon, \quad i = 1, \dots, m$	

The formulations (2.7) and (2.8) above are referred to as the weights formulation of the DEA model. u_r and v_i are the weights attached to the inputs and outputs, respectively, and these are the variables of the model. The input and output weights at the optimal solution can be used to indicate the relative importance of the inputs and outputs in determining the efficiency level of the DMU. However, as these ‘raw’ weights depend on the scaling of each input and output, ‘virtual’ inputs ($v_i^* x_{ij_0}$) and ‘virtual’ outputs ($u_r^* y_{rj_0}$) are used instead. The virtual inputs and outputs are in fact normalised weights, adding up to one for efficient DMUs, both in terms of inputs and outputs. The symbol $*$ will be used in this thesis to denote the value of a variable at the optimal solution to the model in which it appears.

The mathematical infinitesimal (ε) is used to ensure that all inputs and outputs included in the model are taken into account in the efficiency evaluation. See Ali (1990) for a discussion of the choice of appropriate ε values. In practice terms, however, permitting an epsilon weight to be attached to a factor still leads to the virtual zero weighting of inputs and outputs. This means that the factor is effectively omitted from the assessment. The inclusion of weights restrictions in the DEA assessment is discussed in more detail later in this chapter.

The relative efficiency score for DMU j_0 is given by $e_{j_0}^*$ in (2.7) and $1/h_{j_0}^*$ in (2.8). Due to the CRS assumption, the relative efficiency scores provided by the two models are the same, i.e., $e_{j_0}^* = 1/h_{j_0}^*$, see Charnes *et al.* (1978).

By duality, the models (2.7) and (2.8) can be expressed in an envelopment formulation as (2.9) and (2.10), respectively.

DEA input oriented ‘envelopment’ model under CRS	(2.9)
$\text{Min } e_{j_0} = \theta_0 - \varepsilon \left(\sum_{i=1}^m s_i + \sum_{r=1}^s s_r \right)$ $\text{s.t. } \theta_0 x_{ij_0} - \sum_{j=1}^n \lambda_j x_{ij} - s_i = 0, \quad i = 1, \dots, m$ $\sum_{j=1}^n \lambda_j y_{rj} - s_r = y_{rj_0}, \quad r = 1, \dots, s$ $\lambda_j, s_i, s_r \geq 0, \quad \forall j, i, r$	

DEA output oriented ‘envelopment’ model under CRS	(2.10)
$\text{Max } h_{j_0} = \delta_0 + \varepsilon \left(\sum_{i=1}^m s_i + \sum_{r=1}^s s_r \right)$ $\text{s.t. } \sum_{j=1}^n \lambda_j x_{ij} + s_i = x_{ij_0}, \quad i = 1, \dots, m$ $\delta_0 y_{rj_0} - \sum_{j=1}^n \lambda_j y_{rj} + s_r = 0, \quad r = 1, \dots, s$ $\lambda_j, s_i, s_r \geq 0, \quad \forall j, i, r$	

For each assessed DMU j_0 the solution of models (2.9) and (2.10) above seeks to identify a comparator, i.e., a composite DMU corresponding to a linear combination of efficient DMUs $\left(\sum_{j=1}^n \lambda_j^* x_{ij}, \sum_{j=1}^n \lambda_j^* y_{rj} \right)$, with $i = 1, \dots, m$, and $r = 1, \dots, s$, that dominates

DMU j_0 in all input and output dimensions. If $\lambda_j^* > 0$, then the corresponding DMU j is a peer to DMU j_0 .

Additional information obtained from the linear program relates to the slacks on individual constraints (s_i and s_r). The slack variables indicate the extent to which individual inputs or outputs could be improved over and above the amount indicated by the efficiency score. Clearly, at least one input and one output of each DMU j_0 will have a zero slack at the optimal solution. If $s_i^* > 0$ or $s_r^* > 0$ for some i or r , then the DMU j_0 (or its projection on the frontier) lies on an inefficient segment of the frontier of the PPS.

At this point it is useful to distinguish between a radially efficient DMU and a truly efficient DMU in Koopmans' sense.

The radial efficiency of DMU j_0 is defined as follows:

- Input orientation: Radial efficiency is the minimal factor by which all inputs of DMU j_0 can be decreased equiproportionally within the PPS, without decreasing the level of any outputs. With reference to model (2.9), the radial efficiency is θ_0^* . A DMU is radially efficient if $\theta_0^* = 1$.
- Output orientation: Radial efficiency is the inverse of the maximum factor by which all outputs of DMU j_0 can be raised equiproportionally within the PPS, without increasing the level of any inputs. With reference to (2.10), the radial efficiency is $1/\delta_0^*$. A DMU is radially efficient if $1/\delta_0^* = 1$.

A DMU j_0 is efficient (in Koopmans' sense) if and only if the following conditions are both satisfied:

- It has a radial efficiency score of 1;
- It has no positive slack values, e.g., $s_i^* = s_r^* = 0$, $\forall i, r$.

For the inefficient DMUs it is also possible to obtain as by-products of the DEA efficiency assessment a set of targets for becoming efficient. The input and output targets for a DMU j_0 under assessment are obtained as follows:

Targets with an input orientation, from model (2.9):	(2.11)
$x_{ij_0}^{IT} = \theta_o^* x_{ij_0} - s_i^* = \sum_{j=1}^n \lambda_j^* x_{ij} ;$ $y_{rj_0}^{IT} = y_{rj_0} + s_r^* = \sum_{j=1}^n \lambda_j^* y_{rj} .$	

Targets with an output orientation, from model (2.10):	(2.12)
$x_{ij_0}^{OT} = x_{ij_0} - s_i^* = \sum_{j=1}^n \lambda_j^* x_{ij} ;$ $y_{rj_0}^{OT} = \delta_o^* y_{rj_0} + s_r^* = \sum_{j=1}^n \lambda_j^* y_{rj} .$	

The efficiency notion described so far is under constant returns to scale and is referred to as Technical Efficiency (TE). The next section will introduce efficiency measurement under variable returns to scale.

2.4.3.3 The DEA model with variable returns to scale

The returns to scale is a characteristic of the boundary of the technology of production and measures the responsiveness of output to equal proportional changes in all inputs. It assumes that the input mix remains the same whilst the scale size is changed.

The concept of returns to scale can be generalised to the case of multiple inputs and multiple outputs (see Banker *et al.*, 1984). It assumes that the input and output mixes are kept unchanged, and can be expressed as follows:

- A DMU exhibits Increasing Returns to Scale (IRS) if a proportional increase (decrease) in the inputs causes a greater than proportionate increase (decrease) to the outputs.
- A DMU exhibits Decreasing Returns to Scale (DRS) if a proportional increase (decrease) in the inputs causes a less than proportionate increase (decrease) in the outputs.
- Constant Returns to Scale (CRS) are present when a change in the inputs causes an equally proportionate change in the outputs.

These concepts are illustrated in Figure 2.5 for a production frontier defined using DEA.

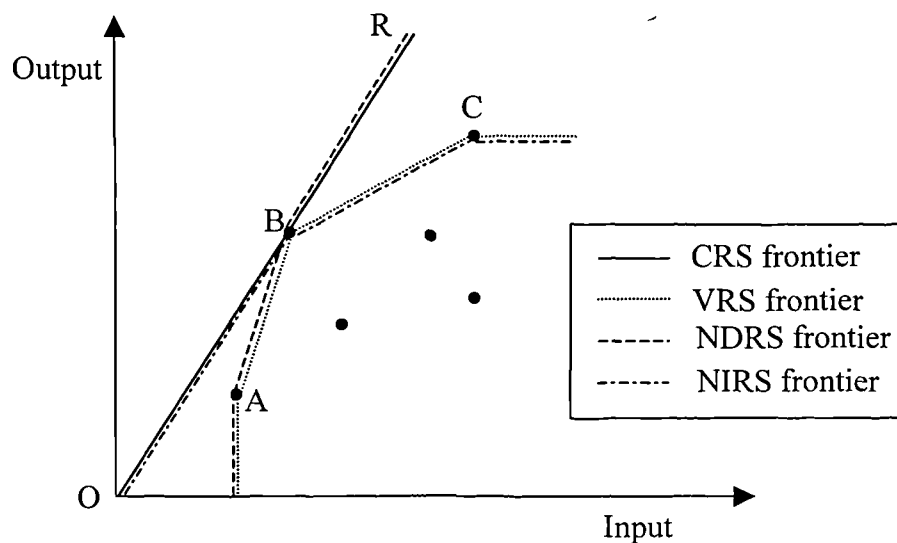


Figure 2.5 – CRS versus VRS frontiers

Under the assumption of CRS, DMU B can be extrapolated to points on the ray OR, such that the change in the input level causes an equally proportional change to the output level. Thus, the CRS frontier is defined by the ray OR.

If the scale extrapolation assumption used in the construction of the CRS frontier is not allowed, the frontier of the PPS must be based on the observed performance of the DMUs given their scale of operation. The efficient frontier in Figure 2.5 would be redefined as the segments between A, B and C. This frontier allows for variable returns to scale and is made of convex combinations of the extreme points lying on the production surface.

Finally, a frontier of mixed character can be developed where extrapolation is permitted for only a subset of efficient DMUs. Let us consider the frontier defined by the segments between O, B and C. This is defined as a *non-increasing returns to scale (NIRS) frontier*. Under this assumption, the scale size of the DMUs can be extrapolated for smaller values, although extrapolations for larger scale sizes are not permitted. It is also possible to define a *non-decreasing returns to scale (NDRS) frontier*, represented in Figure 2.5 by the segments linking A, B and R. Note that it is not possible to specify a DRS or an IRS frontier, as there will always be at least one point on the frontier which has constant returns to scale.

Banker *et al.* (1984) extended the original DEA model to enable the estimation of efficiency under a variable returns to scale context. The VRS models with input and output orientations are provided in (2.13) and (2.14), respectively.

DEA input orientated 'weights' model under VRS	(2.13)
$\text{Max } \hat{e}_{j_0} = \sum_{r=1}^s u_r y_{rj_0} + \omega$ $\text{s.t. } \sum_{i=1}^m v_i x_{ij_0} = 1$ $\sum_{r=1}^s u_r y_{rj} - \sum_{i=1}^m v_i x_{ij} + \omega \leq 0, \quad j = 1, \dots, n$ $u_r, v_i \geq \varepsilon, \quad \forall r, i$ $\omega \text{ is free}$	(2.13a)
For non-increasing returns to scale change (2.13a) to:	
$\omega \leq 0$	
For non-decreasing returns to scale change (2.13a) to:	
$\omega \geq 0$	

DEA output orientated 'weights' model under VRS	(2.14)
$\text{Min } \hat{h}_{j_0} = \sum_{i=1}^m v_i x_{ij_0} + \varpi$ $\text{s.t. } \sum_{r=1}^s u_r y_{rj_0} = 1$ $-\sum_{r=1}^s u_r y_{rj} + \sum_{i=1}^m v_i x_{ij} + \varpi \geq 0, \quad j = 1, \dots, n$ $u_r, v_i \geq \epsilon, \quad \forall r, i$ $\varpi \text{ is free}$	
For non-increasing returns to scale change (2.14a) to: $\varpi \geq 0$	
For non-decreasing returns to scale change (2.14a) to: $\varpi \leq 0$	

The efficiency measured against a VRS frontier is called Pure Technical Efficiency (PTE). In general, under the VRS assumption the orientation of the assessment (input or output) affects the facet of the projection and the resulting DMUs' efficiencies may not be the same. Thus, for inefficient DMUs we may have $\hat{e}_{j_0}^* \neq 1/\hat{h}_{j_0}^*$, although the subset of efficient DMUs is the same irrespective of the model orientation.

It is possible to identify the type of returns to scale exhibited at any point on the VRS frontier by looking at the sign of the intersection of a tangent segment with the output axis. In Figure 2.5 the frontier between A and B (excluding B) exhibits IRS, as the tangent to any point on this segment has a negative intersection on the output axis. Conversely, the frontier between B and C (excluding B) exhibits DRS, as the tangent to any point on this segment has a positive intersection on the output axis. However, at point B we can define several tangent segments to the frontier, ranging from positive to zero and negative intersections with the output axis. A point has CRS when it is possible to define a segment tangent to the frontier that goes through the origin (such as DMU B).

In (2.13) and (2.14) the sign of the variables ω and ϖ can be used to ascertain the nature of returns to scale of pure technical efficient DMUs. Table 2.1 shows how to identify the nature of returns to scale following Banker and Thrall (1992) criteria.

Table 2.1 – Criteria for identifying the nature of returns to scale in the ‘weights’ formulation of the DEA model

<i>In input oriented model (2.13)</i>	<i>Type of returns to scale</i>
$\omega^* > 0$ at ALL multiple optimal solutions	Increasing returns to scale
$\omega^* = 0$ at ANY optimal solution	Constant returns to scale
$\omega^* < 0$ at ALL multiple optimal solutions	Decreasing returns to scale
<i>In output oriented model (2.14)</i>	<i>Type of returns to scale</i>
$\varpi^* < 0$ at ALL multiple optimal solutions	Increasing returns to scale
$\varpi^* = 0$ at ANY optimal solution	Constant returns to scale
$\varpi^* > 0$ at ALL multiple optimal solutions	Decreasing returns to scale

The dual models corresponding to the DEA weights formulations (2.13) and (2.14) are reproduced below in (2.15) and (2.16).

DEA input orientated ‘envelopment’ model under VRS	(2.15)
$\text{Min } \hat{e}_{j_0} = \hat{\theta}_0 - \varepsilon \left(\sum_{i=1}^m s_i + \sum_{r=1}^s s_r \right)$ $\text{s.t. } \hat{\theta}_0 x_{ij_0} - \sum_{j=1}^n \lambda_j x_{ij} - s_i = 0, \quad i = 1, \dots, m$ $\sum_{j=1}^n \lambda_j y_{rj} - s_r = y_{rj_0}, \quad r = 1, \dots, s$ $\sum_{j=1}^n \lambda_j = 1$ $\lambda_j, s_i, s_r \geq 0, \quad \forall j, i, r$	(2.15a)
For non-increasing returns to scale change (2.15a) to:	
$\sum_{j=1}^n \lambda_j \leq 1$	
For non-decreasing returns to scale change (2.15a) to:	
$\sum_{j=1}^n \lambda_j \geq 1$	

DEA output orientated 'envelopment' model under VRS		(2.16)
$\text{Max } \hat{h}_{j_0} = \hat{\delta}_0 + \varepsilon \left(\sum_{i=1}^m s_i + \sum_{r=1}^s s_r \right)$		
$\text{s.t. } \sum_{j=1}^n \lambda_j x_{ij} + s_i = x_{ij_0}, \quad i = 1, \dots, m$		
$\hat{\delta}_0 y_{rj_0} - \sum_{j=1}^n \lambda_j y_{rj} + s_r = 0, \quad r = 1, \dots, s$		
$\sum_{j=1}^n \lambda_j = 1$		(2.16a)
$\lambda_j, s_i, s_r \geq 0, \quad \forall j, i, r$		
For non-increasing returns to scale change (2.16a) to:		
$\sum_{j=1}^n \lambda_j \leq 1$		
For non-decreasing returns to scale change (2.16a) to:		
$\sum_{j=1}^n \lambda_j \geq 1$		

In the VRS envelopment formulation, the sum of λ_j , for $j = 1, \dots, n$, is set to one to prohibit extrapolations of scales of operation. For this type of formulation, the nature of returns to scale of a DMU can be identified based on the analysis of the sum of λ_j values in the constant returns to scale models (2.9) or (2.10). Table 2.2 lists the Banker and Thrall (1992) criteria for characterising returns to scale.

Table 2.2 - *Criteria for identifying the nature of returns to scale in the 'envelopment' formulation of the DEA model*

<i>Envelopment models (2.9) and (2.10)</i>	<i>Type of returns to scale</i>
$\sum_{j=1}^n \lambda_j^* < 1$ at ALL multiple optimal solutions	Increasing returns to scale
$\sum_{j=1}^n \lambda_j^* = 1$ at ANY optimal solution	Constant returns to scale
$\sum_{j=1}^n \lambda_j^* > 1$ at ALL multiple optimal solutions	Decreasing returns to scale

Banker, Bardhan and Cooper (1996) and Banker, Chang and Cooper (1996) proposed simplified methods for testing the nature of returns to scale, which avoid determining all alternative optimal solutions to the ‘weights’ model (see criteria in Table 2.1) or to the ‘envelopment’ model (see criteria in Table 2.2), respectively. An alternative method for the identification of the returns to scale nature was proposed by Fare *et al.* (1985). This method is not affected by the existence of multiple optimal solutions and requires solving three DEA models, with CRS, VRS and NIRS. The nature of the returns to scale is identified by comparing the efficiency measure derived from a NIRS technology and a VRS technology. Banker, Chang and Cooper (1996) prove the equivalence of all these methods. For a review of the methods for the analysis of the returns to scale nature of a DMU using DEA models see Kerstens and VandenEeckaut (1999) and Seiford and Zhu (1999).

The major limitation of the assessments allowing for variable returns to scale is that DMUs with extreme scale sizes (either very large or very small) may be classified as efficient due to the lack of comparators with a similar scale size. Also, the VRS frontier will always envelop the data tighter irrespectively of whether variable returns to scale exist, which may result in an increase in estimated efficiency and less discrimination between the DMUs’ performance. In case it is not known *a-priori* if the production technology exhibits CRS or VRS, Banker (1996) proposes the use of hypothesis tests for the scale effects, such that the VRS model may only be used when the scale effects are demonstrated.

2.4.4 Scale efficiency and most productive scale size

The previous section introduced the DEA efficient frontiers under constant returns to scale and variable returns to scale. The corresponding measures of efficiency were called

For each production process there will always be at least one point on the efficient frontier where the productivity is maximised. This is the part of the frontier for which the tangent through the origin has the greatest slope. The production possibilities for which productivity is maximised are said to be operating at the Most Productive Scale Size (MPSS).

Banker (1984) extended the notion of MPSS to the multiple-input multiple-output situation, considering a given mix of inputs and outputs. The definition of MPSS can be formally stated as follows:

- A production possibility $(X, Y) \in \Phi_{\text{VRS}}$ is a MPSS for its input and output mix, if and only if for all production possibilities $(\alpha X, \beta Y) \in \Phi_{\text{VRS}}$ we have $\alpha \geq \beta$.

Banker and Thrall (1992) showed that the MPSS corresponds to the point(s) on the efficient frontier that maximise the productivity (i.e., α/β) for a given input-output mix. Such point(s) correspond to the intersection between the constant and variable returns to scale frontiers.

Banker (1984) also defined the scale factors $\Lambda^I = \sum_{j=1}^n \lambda_j^*$ and $\Lambda^O = \sum_{j=1}^n \lambda_j^*$, corresponding to the optimal values obtained from the input and output oriented models (2.9) and (2.10), respectively. These factors provide a measure of the divergence of the actual scale size from the MPSS for the input-output mix of a given DMU. Based on the optimal solution to the input oriented model (2.9), a MPSS target is obtained as follows (see Banker, 1984):

$$(x_{ij_0}^{\text{MPSS}}, y_{rj_0}^{\text{MPSS}}) = \left(\frac{\theta_0^* x_{ij_0} - s_i^*}{\Lambda^I}, \frac{y_{rj_0} + s_r^*}{\Lambda^I} \right). \quad (2.18)$$

In Figure 2.6 the MPSS is at point B. All points on the frontier that are not operating at the MPSS can be said to be inefficient. This type of inefficiency is known as scale inefficiency. It captures the amount of inefficiency attributable to the fact that production is not taking place at the MPSS.

Scale efficiency is a measure of how much the scale of operation of a DMU impacts on its ability to achieve maximum productivity. Its relation with technical efficiency and pure technical efficiency is as follows:

$$\text{Scale efficiency} = \frac{\text{Technical efficiency}}{\text{Pure technical efficiency}}. \quad (2.19)$$

The concept of scale efficiency can be illustrated pictorially from Figure 2.6. Adopting an input oriented perspective, the scale efficiency of DMU D is defined as MD_{IC}/MD_{IV} . Note that technical efficiency is equal to MD_{IC}/MD and pure technical efficiency is equal to MD_{IV}/MD , and scale efficiency is equal to their ratio, e.g.,

$$\frac{MD_{IC}}{MD} \bigg/ \frac{MD_{IV}}{MD} = \frac{MD_{IC}}{MD_{IV}}.$$

Adopting an output orientation, scale efficiency is equal to

$$ND_{OV}/ND_{OC}.$$

2.4.5 Economic efficiency

The DEA models discussed so far allow the calculation of technical efficiency and its decomposition into pure technical and scale efficiency components, in case the technology exhibits variable returns to scale. This section introduces the models for the measurement of economic efficiency following Farrell (1957) concepts (i.e., for the measurement of cost efficiency and input allocative efficiency). They can be applied to output oriented assessment (i.e., for the measurement of revenue efficiency and output

allocative efficiency) with obvious modifications, so these other models will not be presented here.

We start defining the concepts of cost efficiency and input allocative efficiency:

- Cost efficiency measures the ability to produce current outputs at minimum cost. It is equal to the ray distance from the current location of the DMU within the PPS to the hyperplane yielding the minimum cost of producing current outputs with existing input prices.
- Input allocative efficiency captures the ability of the DMU to choose the right input mix in the light of prices. It is equal to the ray distance from the frontier of the PPS at the current input mix to the hyperplane yielding the minimum cost of producing current outputs with existing input prices.

In order to obtain a measure of cost efficiency, the minimum cost for the production of a DMU's current outputs with existing input prices is obtained solving the following linear problem, as first formulated by Fare *et al.* (1985):

Minimum cost under CRS	(2.20)
$\begin{aligned} &\text{Min } \sum_{i=1}^m p_{ij_0} x_i^0 \\ &\text{s.t. } \sum_{j=1}^n x_{ij} \lambda_j = x_i^0, & i = 1, \dots, m \\ &\quad \sum_{j=1}^n y_{rj} \lambda_j \geq y_{rj_0}, & r = 1, \dots, s \\ &\quad \lambda_j \geq 0, & j = 1, \dots, n \\ &\quad x_i^0 \geq 0, & i = 1, \dots, m \end{aligned}$	

In the formulation above, p_{ij_0} is the price of input i for the DMU j_0 under assessment.

x_i^0 is a variable that, at the optimal solution, gives the amount of input i to be employed by DMU j_0 in order to produce the current outputs at minimal cost, subject to the technological restrictions imposed by the existing production possibility set.

Note that this model assumes that the price data for each DMU is fixed and known, although the prices may vary from DMU to DMU. Cost efficiency is then obtained as the ratio of minimum cost with current prices to the current cost at DMU j_0 , as follows¹²:

$$\text{Cost efficiency}_{j_0} = \frac{\sum_{i=1}^m p_{ij_0} x_i^{0*}}{\sum_{i=1}^m p_{ij_0} x_{ij_0}}. \quad (2.21)$$

Alternatively, the measure of CE can be obtained with the inclusion of weight restrictions in the standard DEA ‘weights’ model (e.g., model (2.7)). As only relative input prices are relevant for CE measurement, the restrictions imposed to the weights underlying the assessment are that the relative value of the input weights must be equal to the relative value of input prices observed at each DMU, such that:

$$\frac{v_{i^a}}{v_{i^b}} = \frac{p_{i^a j_0}}{p_{i^b j_0}}. \quad (2.22)$$

Where v_{i^a} and v_{i^b} are the weights underlying the cost efficiency assessment and $p_{i^a j_0}$ and $p_{i^b j_0}$ are the input prices observed at DMU j_0 , for any two inputs i^a and i^b used in the assessment.

The resulting cost efficiency model based on the standard DEA formulation with the addition of weights restrictions is as follows:

¹² Note that if all DMUs had similar prices, the cost efficiency measure could be computed by collapsing all inputs to a single input, representing total cost, and solving model (2.7) or (2.9). However, if this procedure is adopted, it is not possible to know the cost efficient targets in terms of individual inputs, which could provide valuable information regarding input mix improvements.

Cost efficiency model	(2.23)
$\text{Max } \kappa_{j_0} = \sum_{r=1}^s u_r y_{rj_0}$ $\text{s.t. } \sum_{i=1}^m v_i x_{ij_0} = 1$ $\sum_{r=1}^s u_r y_{rj} - \sum_{i=1}^m v_i x_{ij} \leq 0, \quad j = 1, \dots, n$ $v_{i^a} - \frac{p_{i^a j_0}}{p_{i^b j_0}} v_{i^b} = 0, \quad i^a < i^b, \quad i^a, i^b = 1, \dots, m$ $u_r \geq \varepsilon, \quad r = 1, \dots, s$	

It can easily be shown that the efficiency measure ($\kappa_{j_0}^*$) obtained from (2.23) is equal to the cost efficiency measure obtained from (2.21), see Schaffnit *et al.* (1997). Note that in the model above, the frontier against which the DMUs are assessed corresponds to a ‘value frontier’ and no longer coincides with the frontier of the PPS defined from the postulates in section 2.4.1. For example, considering the pictorial illustration of Farrell (1957) efficiency concepts in Figure 2.2, the ‘value frontier’ against which the cost efficiency of DMU F would be measured using model (2.23) is defined by the line $P_\alpha P'_\alpha$.

The cost efficiency measure indicates by how much the observed cost could be reduced while being able to secure the observed output. The excess of cost must logically be either because of excess usage of inputs (i.e., technical inefficiency) and/or because inputs are used in the wrong mix in the light of prices (i.e., input allocative inefficiency). The relation between cost efficiency, input technical efficiency and input allocative efficiency is as follows:

$$\text{Cost efficiency} = \text{Input technical efficiency} \times \text{Input allocative efficiency}.$$

As a result, in the DEA framework, the measure of input allocative efficiency can be obtained residually as the ratio of cost efficiency and the input oriented technical efficiency measure.

The efficiency measures outlined in the previous sections are graphically summarised in Figure 2.7, for an input oriented perspective. If input-output quantity data exists, it is possible to measure technical efficiency. This measure can be decomposed into pure technical efficiency and scale efficiency in case the production technology exhibits variable returns to scale. Cost and allocative efficiency measurement is only possible when input prices are also available.

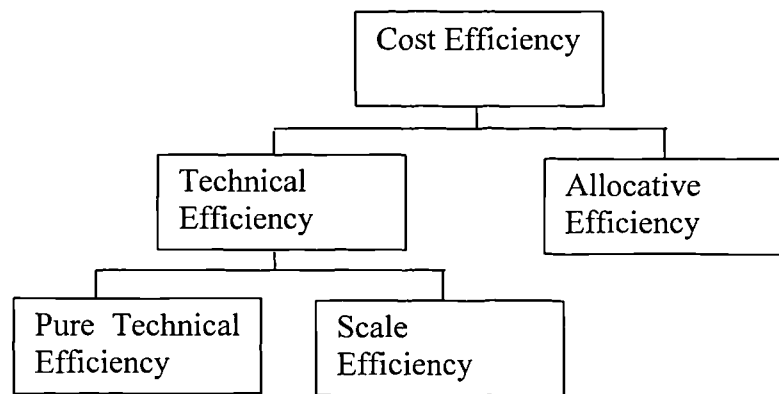


Figure 2.7 – *Efficiency components*

2.4.6 Recent developments

Since its original formulation, DEA has seen considerable expansion during the last twenty years. For a brief synopsis of the evolution of DEA and current state-of-the-art see Seiford (1996). This bibliography reported over 700 papers, which indicates the widespread theoretical and applied expansion of the field.

This section reports a selection of the main extensions and research streams in the DEA literature. Particular emphasis is given to the areas considered of greater importance to the developments in this thesis. The research topics are discussed in the context of the general process underlying a DEA assessment, as illustrated in Figure 2.8. As any other

performance measure, DEA is a cyclical procedure, with inter-linking stages and a feedback mechanism. Figure 2.8 is only a simplified version of the DEA procedure, where the feedback loops are omitted.

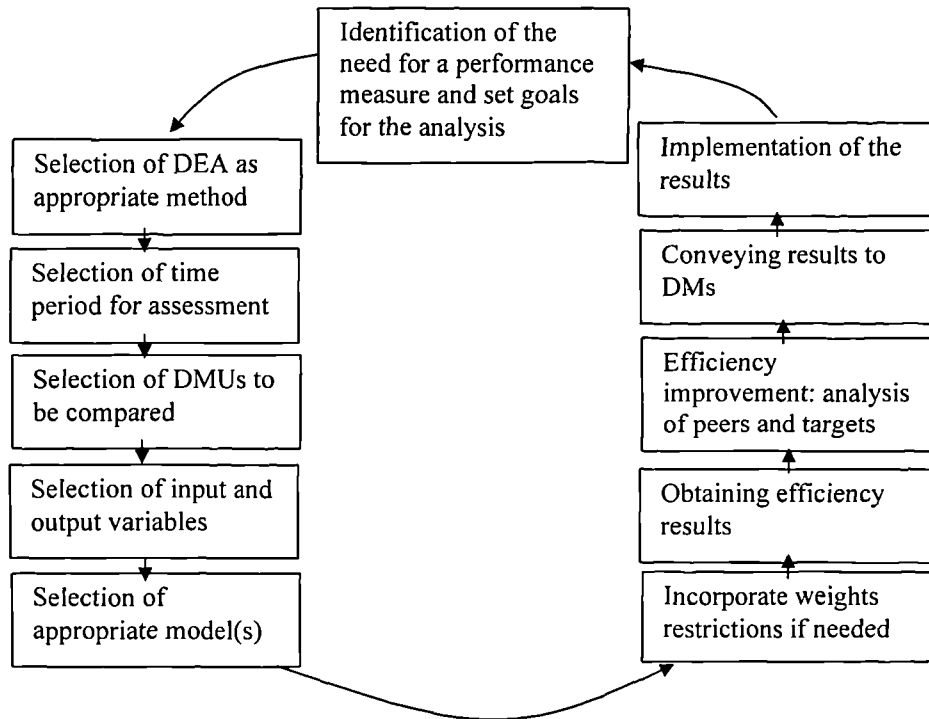


Figure 2.8 – *The DEA application procedure*

Each of the phases of the procedure above are described next, introducing the associated developments and main research agendas.

2.4.6.1 Need of a performance measure and goals of the analysis

In recent years, enormous attention has been devoted to the assessment of performance in organisations. In not-for-profit organisations this was mainly due to the increasing importance of management accountability, which involves reporting not only financial information but also the value of the services provided to the society. In for-profit organisations it is argued that the search for better performance is the only way of assuring competitive advantage and long-run viability. This motivated the development of the efficiency and productivity literature to meet the needs of today's organisations.

The primary aim of performance measurement in organisations remains the derivation of a measure of relative efficiency. At its heart, frontier analysis methods, as discussed in this thesis, are essentially a sophisticated way to benchmark the relative performance of DMUs. They provide an overall, objectively determined, numerical efficiency value that would not be available to organisations otherwise. This ability to quantify issues that decision makers' might only know in a general, qualitative way, makes frontier analysis methods particularly valuable. They allow the identification of areas of input overuse and/or output underproduction, and relate these results to questions of managerial interest.

As performance measurement is becoming a mature and fully established research area, the perspective of the performance assessments is widening from the original aim of providing a summary measure of efficiency. For example, in the context of financial institutions, performance measures have addressed issues such as the effects of deregulation, the consequences of mergers, institutions failure prediction, comparisons of efficiency across international borders and managerial performance improvement, to cite but a few.

2.4.6.2 Selection of the DEA method

In relation to the choice of DEA versus other competing frontier methods, there is a lack of agreement among researchers regarding a preferred frontier model. All methods have their own strengths and weaknesses and a choice of the 'best' method depends very much on the type of data available and objectives of the assessment. For further details on this topic, published studies comparing DEA with other methods include Banker *et al.* (1986), Banker *et al.* (1988), Banker *et al.* (1993), Banker and Cooper (1994), Banker *et al.* (1996), Ferrier and Lovell (1990) and Gong and Sickers (1992).

DEA was originally intended for use in public sector and not-for-profit settings, where typical economic objectives, such as cost minimisation or profit maximisation, may not apply. One of the strengths of DEA is that it can be used even when prices are not available and conventional performance measurement methods are not applicable. However, the application areas of DEA have expanded significantly over the years, and DEA is now used in a wide range of application areas, both in for-profit and not-for-profit organisations.

2.4.6.3 Selection of time period for assessment

The original applications of DEA were based on cross section observations and therefore the efficiency of DMUs was assessed for a particular time period. Charnes *et al.* (1985) introduced the notion of window analysis for assessing performance over time. This was the first use of DEA with panel data. In window analysis, DMUs with data over a number of time periods t are assessed by considering observations from s ($s < t$) adjacent time periods on one cluster. These clusters are treated as cross-sections and DEA is applied consecutively by removing the latest time period from the previous cluster and adding observations from one further period. However, this is an *ad hoc* method, and the number of time periods included in each cluster has to be decided arbitrarily.

The major integration of time into the assessment of efficiency and productivity was later developed by Fare, Grosskopf, Lindgren and Roos (1994) who introduced the Malmquist index approach. This index allows the computation of productivity change and its decomposition into efficiency change and technological change. For a review of the literature on the theoretical developments and applications of the Malmquist index see Fare *et al.* (1998).

2.4.6.4 Selection of DMUs

A DEA assessment requires a set of homogeneous DMUs undertaking similar activities. It assumes that the DMUs use similar resources to produce similar outputs. The DMUs should ideally operate under a common production technology and within similar environments.

However, the above conditions are often not met in real applications. In such cases, one way of improving the homogeneity of the DMUs is to perform cluster analysis prior to the DEA assessment. Once the DMUs have been successfully separated into clusters, the DEA model can be applied to each cluster, with different inputs and outputs, if necessary. See Athanassopoulos (1998) for further details on this topic.

In case the operating environments are not identical, Banker and Morey (1986) modified the original DEA model to allow for the fact that certain inputs may be exogenously fixed and beyond managerial control. This results in a fair comparison of DMUs even when their operating environments are not homogeneous.

In some cases, the objective of the assessment may be the evaluation of the differences in DMUs' performance caused by operating under different programs or environmental conditions. The most significant break-through for this purpose was the method developed by Charnes *et al.* (1981) in the context of the evaluation of the Program Follow-Through versus Non-Follow-Through in the US. This procedure can be enhanced using hypothesis tests, as illustrated in Ward *et al.* (1997) and Elyasiani and Mehdiian (1995).

2.4.6.5 Selection of input-output variables

In order to represent the DMUs' activity as accurately as possible, the input-output set should cover the full range of resources used and capture the outputs that are relevant for

the objectives of the analysis. However, if large numbers of factors are used in the analysis, the ability of DEA to distinguish between the efficient and inefficient DMUs decreases. Therefore, the choice of inputs and outputs should be parsimonious to achieve discrimination. See Banker *et al.* (1989) and Dyson *et al.* (1990) for a suggestion of rules of thumb on the maximal number of variables to be included in the assessment¹³.

The input-output variables defined should be such that it is desired to minimise input levels and maximise output levels. However, undesirable outputs, such as a pollutant, do not conform to the above criteria, as increasing the level of this output should reduce the efficiency level of the DMU. This type of situation is dealt with in the literature by inverting the factor, subtracting the value from a large number, or moving the output variable to the input side of the model. However, these methods can lead to different efficiency results, and the appropriate method to deal with this issue is still not clear.

The input-output variables may be qualitative or quantitative in nature. When inputs and outputs are not easily measurable or are not available, proxy variables may be chosen to represent them. However, their inclusion in the DEA model requires that the values conform to ratio scales (i.e. there is a meaningful zero and equal intervals have equal value). In case the variables are not measured in a ratio scale, Banker and Morey (1986) developed a DEA model that can handle categorical data. For models that can deal with ordinal data see Cook *et al.* (1993) and Cook *et al.* (1996).

An established methodology for selecting the factors of a DEA assessment can be found in Golany and Roll (1989). For a discussion of problems associated with variable

¹³ Banker *et al.* (1989) suggest that number of DMUs should be greater than 3 times the sum of the number of inputs and output. Dyson *et al.* (1990) suggest that the number of DMUs should be considerably greater (i.e., two or three times greater) than the product of the number of inputs and number of outputs.

selection and measurement see Nunamaker (1985), Sexton *et al.* (1986), Smith (1997) and Stolp (1990).

2.4.6.6 Selection of the DEA model(s)

Since the development of the original DEA model, the literature has been rapidly expanding and a large number of alternative models are now available. These models can be classified according to:

- Orientation;
- Assumptions about returns to scale;
- Disposability of inputs;
- Type of efficiency measure calculated;
- Convexity assumption relating to the shape of the frontier.

In terms of orientation, the models may have an input orientation, an output orientation, or both. In the latter case, they are called graph efficiency models (see Fare *et al.*, 1985). Recently the literature has also proposed the use of directional efficiency measures (Chambers *et al.*, 1996).

In terms of returns to scale assumptions, the models can handle CRS, VRS, NIRS or NDRS. In case the returns to scale underlying the production technology are not known *a-priori*, the choice of the appropriate type of returns to scale model can be made following the procedure suggested by Banker (1996).

The DEA model can either assume strong input disposability¹⁴ or weak input disposability¹⁵ (see Fare, Grosskopf and Lovell, 1994).

¹⁴ Strong (or free) input disposability implies that if any input increases then output does not decrease.

¹⁵ Weak input disposability implies that if all inputs increase proportionally the output does not decrease. This property allows for the possibility that one input is 'bad' and thus its increase beyond certain levels (causing congestion) may result in a decrease in output, keeping the other inputs constant. Weak input disposability is used to model situations where backward bending isoquants are feasible.

Another distinction between the models refers to the resulting efficiency measure, i.e. radial versus non-radial. The latter account for slacks in inputs and outputs and thus satisfy the Koopmans (1957) definition of efficiency. Fare and Lovell (1978), Zieschang (1984), Fare *et al.* (1983), Pastor *et al.* (1999) and Cooper *et al.* (1999) proposed alternative non-radial efficiency measures.

Finally the convexity assumption of the original DEA model can be relaxed leading to the Free Disposal Hull model (Deprins *et al.*, 1984) or to the non-convex formulation proposed by Petersen (1990).

2.4.6.7 Weights restrictions

The original DEA model as developed by Charnes *et al.* (1978) was based on the assumption that each DMU should have free choice in selecting weights for inputs and outputs. This era lasted until 1986, when Thompson *et al.* (1986) argued that in selecting potential sites for the location of a nuclear research laboratory, they had to restrict the flexibility of the weights to enable discrimination between technically efficient DMUs. This was followed by a rapid extension of the literature in this area, in order to incorporate judgement in the DEA models through the restriction of multipliers, see Dyson and Thanassoulis (1988), Wong and Beasley (1990), Charnes *et al.* (1990), Thompson and Thrall (1994), Thanassoulis *et al.* (1995). For a review of the literature on weights restrictions see Allen *et al.* (1997).

2.4.6.8 Efficiency results

The DEA model is solved using linear programming techniques. As DEA requires the solution of separate linear programmes for each assessed DMU, this requires the generation of a sequence of similar but not identical problems to be solved. In recent years, the choice of appropriate software has become increasingly important for the

large-scale and complex DEA studies being conducted. It is either possible to use specialised DEA software (e.g. Warwick DEA, Frontier Analyst, IDEAS) or a standard LP package (e.g., AIMMS, GAMS, SAS, LINDO, EXCEL, XPRESS-MP, GAUSS, MATLAB, MATHEMATICA, CPLEX). The empirical analysis reported in this thesis used the Warwick DEA software and AIMMS.

2.4.6.9 Efficiency improvement: Peers and Targets

The DEA models have many by-products beyond efficiency measures with profound managerial implications. The efficiency assessment is based on the comparison of inefficient DMUs to points on the efficient frontier. Thus, each inefficient DMU can be given a set of input-output targets that would render it efficient, obtained from its projection to a point on the frontier. In some studies, performance improvement issues have been recognised as potentially more important in the long-term than mere measures of efficiency.

The derivation of efficient targets for inefficient *DMUs* should be analysed independently from the efficiency measure. For example, in CRS models the input and output oriented efficiencies are equal. However, the corresponding targets yield different points on the frontier. Some researchers have specifically addressed target setting issues in a DEA framework. Further details on this topic can be found in Thanassoulis and Dyson (1992), Kao (1994), Zhu (1996) and Athanassopoulos (1994, 1995).

2.4.6.10 Conveying results to decision makers

Obtaining meaningful performance evaluation results is certainly an important step in the DEA analysis. The ability to explain them across the organisation is also vital in order to validate the results and ensure proper implementation of the findings. However, this is still an area that needs further development within the DEA literature. As reported in a

Fortune article by Norton (1994), one reason for the slow migration of DEA to the business world is the ‘black box’ syndrome, i.e., the reliance on complex mathematics. In this Fortune article, an economic systems director recalled a presentation in which the clients were impressed and ready to use DEA analysis to help determine incentive pay – until a senior manager vetoed the idea with a simple assertion: “Our store managers are never going to understand it”. Research on this important issue of how to communicate results to managers can be found in Belton and Vickers (1993) and Schaffnit and Paradi (1998).

2.4.6.11 Implementation of the results

This is still an issue that has not often been reported in the literature. To date, only few studies have noted in any detail the specific changes implemented to improve performance at inefficient DMUs (see Sherman and Ladino, 1995).

2.5 Summary and conclusions

This chapter presented the frontier analysis methods for the evaluation of efficiency in organisations. It emphasised the DEA method, which will be used throughout the thesis. The theory underlying the construction of the production possibility set and the efficient frontier of a DEA assessment was described. The basic DEA models were reviewed and the associated efficiency estimates described (i.e., cost efficiency, input allocative efficiency, technical efficiency, pure technical efficiency and scale efficiency). The recent developments in the DEA literature were presented.

The next chapter focuses on the use of the methods and models described in this chapter for the analysis of efficiency in the banking sector. It reviews the past research in this area and connects the current state-of-the-art with the themes and questions addressed in this thesis.

CHAPTER 3

The analysis of efficiency in the banking sector: An overview of past research and future directions

3.1 Introduction

This chapter reviews the literature on banking efficiency measurement. It describes the state-of-the-art on the assessment of performance of banks and bank branches, and highlights the directions in which future research might be most fruitful. The purpose of this review is to study and summarise the main aims, methodologies and conclusions of previous research on banks' efficiency. The information gathered will be used in the empirical part of this thesis, enabling the discussion of the empirical results in the light of previous studies and the generalisation of some conclusions.

Recent developments in financial markets, such as liberalisation, globalisation, and technological evolution, have contributed to a substantial increase in competition among financial institutions. These changes have put an increasingly sharp focus on the analysis of banks' activity, as achieving high performance standards became essential to retain competitiveness and assure corporate health. Therefore, the topic of financial institutions' efficiency has recently attracted widespread attention from bankers, regulators and the scientific community. The number of studies has increased substantially, alongside the number of journal special issues dedicated to the assessment

of performance of financial institutions¹. For a recent survey of studies that applied frontier methods to the analysis of financial institutions' efficiency see Berger and Humphrey (1997).

The performance measurement literature related to the banking sector has been predominantly concerned with the assessment of corporate performance, based on comparisons of efficiency across banks. However, the improvement of performance in each individual bank is very much dependent on effective branch management. Consequently, the number of studies focusing on the relative efficiency of branches within a particular bank is increasing. Unfortunately, the data required for this type of study is not easily available, which is possibly one of the main causes for the still relatively scarce literature on this important topic.

This chapter is structured as follows. Section 3.2 discusses the comparative advantages and limitations of 'traditional' performance measures versus frontier methods applied to the banking industry. Section 3.3 focuses on the definition of banks' activity and the selection of the inputs and outputs for an efficiency assessment. Section 3.4 summarises the aims, methodologies and conclusions of frontier studies on banking efficiency, with particular emphasis given to the studies based on the Data Envelopment Analysis method. Section 3.5 summarises and concludes.

¹ E.g., the Journal of Productivity Analysis (1993), the Journal of Banking and Finance (1993), the European Journal of Operational Research (1997) and Interfaces (1999).

3.2 Alternative methods for the assessment of performance in banking

This section examines the alternative methodologies to assess performance in the banking sector. Ratio analysis and profitability measures have for long been used before the development of the frontier approaches outlined in the previous chapter. The relative strengths and weaknesses of these methods are discussed next.

3.2.1 Ratio analysis

Ratio analysis has traditionally been the method of choice in assessments of corporate performance. It typically involves the use of a series of input and output ratios that can be classified in four major categories: leverage ratios, liquidity ratios, profitability ratios and market value ratios (Brealey and Myers, 1996).

Though often criticised, in practice the use of ratios continues to be the preferred method for assessing performance in many institutions. A few factors can contribute to this popularity. Firstly, some of these measures are demanded by regulatory agencies. E.g., in the financial sector bank regulators often use financial ratios of accounting data to screen banks. Also, their familiarity, interpretability, and ease of calculation contribute to their continued widespread use.

In spite of the generalised acceptance of ratio measures as performance indicators, there are a few problems associated with their use.

Firstly, there is an implicit assumption of constant return to scale in financial ratios. In order to enable a comparison of DMUs allowing for their scale size, the construction of the ratios assumes proportionality between the numerator and denominator.

Secondly, each individual ratio measures performance in relation to one input and one output only, and thus examines only a part of the DMUs' activity. To overcome this limitation, the performance evaluation is generally based on the analysis of a set of ratios. However, it is still difficult to gain an overall view of performance, as the number of performance indicators that can be computed for each DMU may become unmanageably large. The ranking of DMUs is also impaired, particularly when not all ratios indicate a similar level of performance for each DMU.

Aggregating the multiple dimensions of the DMUs' activity evaluated by each ratio into a single summary measure of performance is frequently necessary. But while the calculation of a set of ratios is a relatively easy task, the aggregation of those ratios can be a quite complicated process, involving imagination and experienced judgement. The literature describes several methods to combine ratios by various weighting systems. However, the measure resulting from the aggregation of a set of ratios can bias the view on performance that the individual ratios convey.

Another limitation of ratios is that they cannot be used in a straightforward manner to set performance targets. This is because each single ratio has to be compared to some benchmark value, without regarding the remaining input-output levels of the DMU concerned, and assuming that the benchmark chosen is suitable for comparison. Although any particularly poor value of a ratio identifies an aspect of the activity in special need for improvement, the target performance levels cannot be estimated with confidence, as a target for one ratio may have implications for others

In addition, ratio measures can not explicitly account for the input-output mix of the DMUs. To illustrate this point, consider the ratio of cost per teller transaction, which is often used in bank branches to evaluate operating efficiency. The idea underlying the use

of this ratio is that the branches with the highest costs per transaction would be potentially the least efficient. However, a higher cost per teller transaction may be due to a more complex mix of transactions, which is not accounted for in the construction of the ratio, (e.g., a branch that primarily opens new accounts and sells credit to customers would require more resources than a branch that primarily processes less complex transactions, such as deposits and check cashing).

In an attempt to overcome some of the limitations of the use of ratio methods for performance assessment, researchers have studied their relation to more modern and powerful approaches such as frontier methods. Elyasiani *et al.* (1994) and Yeh (1996) investigated the relationship between a bank's financial performance, measured by accounting-based ratios, and efficiency levels obtained using DEA. Smith (1990) applied DEA to financial statements. The method described can be used to dissect the information provided by a traditional ratio in order to identify the main issues affecting the companies' performance. Fernandez-Castro and Smith (1994) used DEA to explore DMUs' performance based on financial ratios. Using the financial ratios as the inputs and outputs of a DEA model can overcome some of the limitations of conventional ratio analysis, e.g., the identification of suitable peers and targets for inefficient DMUs. Thanassoulis *et al.* (1996) compared DEA and ratio analysis, drawing conclusions about their relative strengths and weaknesses from a generic perspective.

3.2.2 Profitability

It is often argued that in the corporate sector performance is ultimately judged on the basis of long-term profitability. The indicators of profitability most frequently used are ratios of data reported in financial statements. A company is then evaluated comparing

its ratios with some standard value, such as a location measure for the industry (i.e., the mean or median).

Although profitability is a highly relevant measure of performance, companies are nowadays complex organisations with a multivariate nature, such that judging performance against a single yardstick such as profitability will generally result in a partial and imperfect assessment.

Also, profitability measures do not reflect the ability to use all resources to the maximum advantage, as they ignore efficiency issues in carrying out the business activity. In the context of banking business, higher profits may be associated with more favourable market environments. For example, the profitability of a bank branch does not indicate whether the resources are being used efficiently. A branch that processes a high proportion of cash withdrawals and other services which do not generate revenue may have lower profitability than one that processes a higher proportion of revenue generating transactions. Nevertheless, the less profitable branch may be more efficient in using its personnel and other resources to provide the customer services required.

Also, profitability indicators have inevitably a short-term nature. They reflect the current achievements and fail to take into account the value of strategic actions and investment decisions that will affect future as opposed to current performance. For example, a bank that defers marketing or new product development costs may appear to be performing well based on financial ratios, even though these actions may impair future performance.

Kaplan and Norton (1992) developed the *balanced scorecard* to address the inability of traditional measurement systems to link a company's long-term strategy with its short-term actions. The *balanced scorecard* supplements traditional financial measures such as profitability with criteria that measures performance from three additional perspectives -

those of customers, internal business processes, and innovation and learning. This methodology enables companies to track financial results while simultaneously monitoring progress in building the capabilities and acquiring intangible assets they would need for future growth². However, despite being an important complement for financial measures, the *balanced scorecard* has yet to gain widespread acceptance in the banking sector.

The next section discusses the use of frontier methods to analyse financial institutions' performance.

3.2.3 Frontier methods

Prior to the late 80s, the studies of banking performance were mostly based on financial ratios, sometimes associated with econometric techniques, such as multivariate regression. These studies concentrated on measuring scale and scope economies, based on average performance standards, and implicitly assuming efficiency in banking activity.

The pioneering work on banking efficiency, enabled by the use of frontier methods, dates back to 1985 (Sherman and Gold, 1985). This study evaluated the efficiency of bank branches from the same institution, adopting an operational perspective.

Later research has found that the scale and scope economies, which had been extensively studied in the past, accounted for less than 5% of costs, whereas inefficiency accounted, on average, for around 20% of banking costs (Berger *et al.*, 1993). This explains the

² In a recent article, Rouse *et al.* (1997) discussed how DEA and profitability measures could be incorporated in a general managerial framework for performance assessment, such as the balanced scorecard. They illustrated the framework proposed with an application to a highway maintenance setting.

widespread attention that frontier methods have attracted within the banking sector during the 90s.

Five different types of frontier methods have been frequently employed in the evaluation of financial institutions' efficiency (i.e., Data Envelopment Analysis, Free Disposal Hull, Stochastic Frontier Approach, Distribution-Free Approach and Thick Frontier Approach). These methods differ primarily in how much shape is imposed on the frontier and the distributional assumptions used to disentangle efficiency differences from random errors, as discussed in the previous chapter. The most widely used methods for the assessment of banking efficiency have been the Stochastic Frontier Approach (SFA) and Data Envelopment Analysis (DEA).

The literature has not yet reached a consensus on the preferred method for determining the best-practice frontier against which relative efficiencies should be measured in financial institutions.

The SFA can handle statistical noise, but it imposes an explicit and possibly overly restrictive functional form for technology. If the functional form is misspecified, the inefficiencies identified may be confounded with the specification error. In addition, it requires imposing an explicit distribution for the inefficiency term.

The DFA does not require defining a-priori a distribution for the inefficiency term. However, it requires panel data and assumes the efficiency level does not change over time.

Among the parametric methods, the TFA requires the least demanding assumptions regarding the form of the frontier and the distribution of inefficiency. However, it cannot provide estimates of efficiency for individual DMUs.

In addition, none of the parametric approaches described above (i.e., the SFA, DFA and TFA) can easily provide estimates of allocative efficiency or handle multiple inputs and outputs.

In relation to the non-parametric approaches, both the DEA and FDH approaches impose less structure on the frontier but commit the sin of not allowing for random variations in the data, owing to luck or measurement errors. If random errors exist, efficiency will be confounded with the random deviations from the frontier.

The lack of agreement among researchers regarding a preferred frontier model at present boils down to a difference of opinion regarding the lesser of evils. Although the stronger assumptions of the econometric methods may generate stronger results when the functional form of the frontier and the distribution of inefficiencies can be accurately estimated, minimal assumptions are preferable when this information is not available *a-priori*. The parametric models are often limited by the absence of theoretical foundations and reliable statistical tests for the functional specification of the frontier and the distribution of the inefficiency term, which is not known *a-priori*.

In addition to the greater flexibility of the assumptions imposed on the frontier estimated by DEA, this method can easily handle multiple inputs and outputs and allows a straightforward decomposition of the efficiency estimates into allocative, pure technical and scale efficiency for each individual DMU. As the data used in the empirical part of this thesis was obtained from a bank's internal performance measurement system, the data is likely to be cleaner and more detailed than most banking data sets. Thus, the use of a method such as DEA, requiring minimal assumptions regarding the shape of the production frontier, whose major limitation is not accounting for the existence of random errors, is the most appropriate for the empirical part of the thesis.

As DEA was the chosen method for this thesis, the review of banking efficiency assessments presented in this chapter is particularly focused on the use of the DEA method. Before discussing the aims and conclusions of previous studies, the issues related to the definition of banking activity are described next.

3.3 The inputs and outputs of banking

A fundamental difficulty of the analysis of banking activity is the characterisation of the production process. Banking institutions provide services, rather than readily identifiable physical products, and there is no general consensus regarding the specification of appropriate inputs and outputs. An unsettled controversy remains in the literature, which has given rise to alternative approaches to the definition of banks' activity. This section seeks to identify the conceptual issues regarding the definition of the bank inputs and outputs and describe the main approaches used in empirical studies.

As described by Humphrey (1985), the basic business of financial institutions is the process of intermediating equity capital and liabilities (e.g., deposits from customers and purchased funds from the interbank monetary market) into assets (e.g., loans, securities and other investments). Banks' revenue is generated from the interest payment received from loans and the commissions from other financial services provided to customers (e.g., cross selling activities associated with insurance and investment funds³). Simultaneously, banks pay depositors an interest compensation for the provision of funds. Depositors also receive other non-monetary services, such as liquidity, safekeeping and accounting services. The execution of this process requires the use of real resources such as labour, capital, materials and information systems.

³ The financial services for which banks receive commissions are often referred to as off-balance sheet activities.

The literature contains a variety of approaches to the assessment of efficiency in banking. Considering the characterisation of the banking business and the associated definition of inputs and outputs, the studies can be categorised in five different approaches, referred to as the production, intermediation, asset, user cost and value-added approaches. All these approaches can be used in the analysis of the relative efficiency of banks or for the assessment of branches within a bank.

3.3.1 Production Approach

The production approach emphasise the operational activity of the bank, and thus banks are primarily viewed as providers of services to customers. Under this approach, the inputs include physical variables related to operating costs (e.g., labour, materials, space or information systems) and the outputs are measured by the workload of operational activities. This approach was introduced by Benston (1965) and mainly characterised the literature up to the early 1980s.

In relation to the input set, only physical inputs such as labour and capital or their associated costs should be included, since only physical inputs are needed to perform transactions, process financial documents or provide other types of services to customers. Interest costs are excluded from this approach on the grounds that only the operational process is of interest. Benston *et al.* (1982, p.9) noted that “*while interest is an important outlay to the bank, it is determined by market forces that reflect alternative investments available to depositors. Thus, interest is not an operating expense for purposes of measuring banks’ efficiency*”.

The output of this approach represents the services provided to customers and is best measured by the number and type of transactions or documents processed over a given time period. It is common to group the transactions according to the level of resource

consumption, to their complexity or to their purpose, which can help the interpretation of the efficiency results.

However, such detailed transaction flow data is not generally available, and data on the stock of deposit and loan accounts is often used instead, as a proxy for the level of services provided. In studies focusing on the efficiency of branches within a bank, the use of these proxy variables can be problematic, as an account can be opened at one branch, but transactions processed at other branches. Thus, the workload is not accurately evaluated unless some adjustments for interbranch transactions are incorporated in the model.

In addition, there is a lack of consensus on the output definition as the number and/or value of accounts (whenever data on the number of transactions is not available). Most analysts argue that although the value of the accounts may affect to some extent the operational costs, the number of accounts dealt with determines primarily the operational costs.

The main drawback of using as output the number of accounts is that the banks can have a significant number of so called 'dead accounts', which are not used and almost do not have deposited funds. This situation can arise when the depositors work with two or more banks.

Different arguments have been put forward for using the value of accounts as the output measure: banks compete to increase their market share regarding the monetary value intermediated, as opposed to the number of accounts, and large accounts can be more costly than small accounts since they tend to be more active.

The production approach is the most widely used in the analysis of bank branches' efficiency. One of the reasons why the production approach has rarely been used for efficiency studies at the bank level is the difficulty encountered in collating accurate data. The type of data needed for this approach is not openly available in the majority of countries, as the information required by supervision authorities and published by the banks is mainly financial.

3.3.2 Intermediation Approach

Under the intermediation approach financial institutions are viewed as primarily intermediating funds between savers and investors. Banks produce intermediation services through the collection of deposits and other liabilities and their application in interest-earning assets, such as loans, securities and other investments.

In general, inputs are measured by interest and non-interest costs and outputs reflect the financial flows associated with the intermediation activity. Since intermediation services flow data is not usually available, the flows are typically assumed to be proportional to the stock of financial value in the accounts, as expressed in the balance of the bank.

There is a long-standing controversy whether deposits should be considered inputs or outputs. The argument to include deposits as inputs is that they generate costs and do not produce revenue until they are intermediated into loans and other assets. The funds raised by deposits provide the raw material of investable funds, and it is the ultimate use of these funds that generates the bulk of direct revenue that banks earn.

More recent studies have included deposits as outputs. They are seen as an additional product over which banks compete. In addition, it is recognised nowadays that deposits are resource consuming and contribute to a substantial proportion of banks' costs. Also,

their volume can serve as a proxy for non-monetary services provided to depositors as compensation for the use of their funds. Most studies resolve the issues raised by the inclusion of deposits in the input-output set with an approach that captures both the input and output characteristics of deposits. The interest paid on deposits is included as input, whereas the monetary value of deposits is specified as an output.

Some studies have refined the definition of the input-output set by making distinctions between different types of deposits. For example, Rangan *et al.* (1988) considered demand, term and saving deposits as outputs, while purchase funds were regarded as inputs. This differentiation was considered necessary because the latter are not highly resource consuming.

A few studies have defined outputs in terms of the revenues generated by the intermediation process. This specification of outputs in terms of revenues can be very relevant to the analysis of the economic viability of banks, since the value of the accounts does not indicate whether a bank obtains a high return from the intermediation activity. Using interest revenue as an output incorporates simultaneously two aspects of banking activity in the model, as interest revenues are a function of the *amount* of lending as well as the *price* charged for the loans. However, revenue can be more sensitive to market and environmental conditions than to internal managerial issues, which may be best represented by the accounts' value.

The main shortcoming of the studies adopting the intermediation approach is that most services provided by banks but not proxied by balance sheet magnitudes are rarely taken into account, as argued by Favero and Papi (1995). This omission is particularly relevant nowadays, as the increase in competition has led to a reduction in the intermediation margins of the banks. This has forced banks to obtain revenue from the commissions

charged for the new financial services offered to clients, such as the selling of investment funds, insurances, securities and other brokerage activities. As noted by Sauders (1993), the off-balance sheet activities of the largest US banks often exceed the on-balance activities of these banks by a factor of four or five when measured by monetary values.

The intermediation approach has been the most widely used in the studies of efficiency at the bank level.

3.3.3 Asset Approach

The asset approach is a reduced form modelling of the banking activity, focusing exclusively on the role of banks as financial intermediators between depositors and those that receive bank loans. Deposits and other liabilities, together with real resources (labour and capital) are defined as inputs. Because only bank assets are specified as outputs, this approach is usually termed the asset approach⁴. This approach was first suggested by Sealey and Lindley (1977).

It is argued that bank liabilities have some characteristics of inputs because they provide the raw material of investable funds, and that bank assets have some characteristics of outputs as they generate the bulk of the direct revenue that banks earn.

For those banks that primarily purchase their funds (with interest payments) from other banks and turn these funds into loans, this is an adequate description of bank output. However, most banks do much more than purchase their funds - they also provide substantial services to depositors that are not accounted for in the asset approach. The attraction of customer deposits is nowadays essential to the banking business, such that a

⁴ This approach is sometimes referred to as the intermediation approach, although the flexibility in the choice of inputs and outputs that characterised the intermediation approach is replaced by a rigorous definition of the input and output variables.

significant share of the operational costs can be incurred by activities related to the attraction and maintenance of deposits.

Another limitation of this approach is related to the fact that only the data reported in the balance is included as output. Thus, any type of financial product sold by the bank, such as investment funds or securities, is not accounted for in the output set, despite the growing importance of commissions as a source of banks' revenue.

3.3.4 User Cost Approach

The user cost approach determines whether a financial product is an input or an output on the basis of its net contribution to bank revenue. If the financial returns on an asset exceed the opportunity cost of the funds or if the financial costs of a liability are less than the opportunity cost, then they are considered as outputs. Otherwise, they are considered as inputs. Hancock (1985) was the first to apply the user cost approach to banking.

There are two main criticisms to this approach: the difficulties in collecting accurate data and the practice of subsidisation, which implies low reliability of the available data on costs and revenues. It is also difficult to determine opportunity costs for bank assets and liabilities taking into account important characteristics such as credit risk, liquidity and maturity (see Berger and Humphrey, 1992).

3.3.5 Value-Added Approach

The value-added approach identifies as outputs those balance sheet categories (assets or liabilities) that highly contribute to the bank's value added (i.e., business associated with the consumption of real resources). In general under this approach, the major categories of produced deposits (e.g., demand, term and saving deposits) and loans (e.g., mortgages and commercial loans) are viewed as important outputs because they are responsible for

the great majority of value added. The inputs are labour, capital and purchased funds. The purchase funds are treated as *inputs* because *they require very small amounts of physical inputs*. Finally, off-balance sheet business is often excluded from the output list due to their small contribution to value-added (see Berger and Humphrey, 1992).

3.3.6 Discussion

Although these five approaches to the definition of inputs and outputs are well established, the choice of inputs and outputs of banking efficiency studies is still very much influenced by the analyst's own view of the activity of the bank, the issues under analysis, and the availability of data.

The asset, user-cost and value-added approaches can be viewed as variants of the intermediation approach, corresponding to different trends in the debate on the identification of banking output, as noted by Berger and Humphrey (1992). In fact, these three approaches are focused on the intermediation activity of the banks and mainly use financial data.

The two approaches most frequently used in the assessment of financial institutions are the production and intermediation approaches. However, none of these two approaches is perfect because neither fully captures the dual roles of financial institutions (Berger and Humphrey, 1997): (i) providing transactions/document processing services to customers and (ii) being financial intermediaries that transfer funds from savers to investors at a profit. Thus, the use of both approaches would be advisable in the analysis of financial institutions, whenever there is sufficient data to implement such a research design. Nevertheless, each of the approaches has a specific scope so that the results and conclusions drawn can be particularly valuable for certain purposes.

It is argued that the intermediation approach may be more appropriate for evaluating entire financial institutions and addressing questions concerning the economic viability of banks, as it includes both operational and interest expenses (Ferrier and Lovell, 1990; Berger *et al.*, 1987). The intermediation approach may also be superior for evaluating the importance of efficiency to the profitability of the financial institution, since minimisation of total costs, not just operational costs, is needed to maximise profits (Berger and Humphrey, 1997). As the production approach includes only operating costs and leaves out interest expenses, which account for about 50% to 70% of total bank costs, it should only be used to assess the operational aspects of banking activity, as argued in Berger and Humphrey (1991 and 1997).

The production approach may be better for evaluating the efficiency of branches from a financial institution. The branches primarily process customers' documents and typically have little influence over fund raising and investment decisions. Also, under some circumstances, the economic viability of a branch may be sacrificed in favour of other strategic objectives.

3.4 The measurement of efficiency in banking

This section reviews the aims, methodologies and conclusions of the studies that applied frontier methods to the analysis of financial institutions. The review focuses on studies published in journals. Particular emphasis is given to the applications using DEA, which is the technique used in the empirical part of this thesis.

During the late 80s and particularly in the 90s frontier methods have been used extensively to evaluate banking institutions. Most of these studies were motivated by the desire to measure efficiency and improve performance.

Alongside the growing interest in the financial services sector observed in most countries, some methodological issues related to the measurement of financial institutions' efficiency have been extensively researched. The main topics studied include:

- The impact of frontier specification on efficiency estimates;
- The identification of inefficiency sources (i.e., technical, scale or allocative issues);
- The definition of inputs and outputs and sensitivity of efficiency results to variable specification;
- The choice of the orientation for the assessment (input versus output orientation).

Other studies have addressed issues that provide valuable information relating to regulatory/government policy for the financial sector, such as:

- The effects of deregulation on banking efficiency;
- Prediction of financial institution failure;
- Organisational form and ownership impact on efficiency;
- Mergers impact on efficiency;
- Inter-country comparisons.

The assessment of efficiency is inevitably associated with the desire to improve performance. Perhaps the best potential use of frontier methods in improving the performance of financial institutions comes from efficiency analysis of the branches within an individual institution. These studies can be particularly useful for bankers if detailed proprietary data is available, enabling an accurate representation of the business activity.

The DEA studies reviewed are summarised in table format in the appendix to this chapter. This table contains information about the authors and the year of publication, the inputs and outputs used, the orientation and type of efficiency measures computed, the

methods used, the country and type of data, and the main aims and conclusions. The following paragraphs summarise the general conclusions of this review.

3.4.1 The impact of frontier specification on efficiency estimates

Both parametric and non-parametric methods have been extensively used for banking efficiency measurement. In terms of the assessments at the corporate level, there is a lack of consensus about the most adequate method. In relation to the analysis of efficiency of branches within a parent bank, most studies have used non-parametric methods. This may reflect a consensus among researchers that this method is more appropriate for branch performance measurement. This is probably because this type of study must be based on proprietary data, which is generally cleaner than the published financial data used in most studies at the corporate level.

Berger and Humphrey (1997) compared the efficiency values obtained using these two broad classes for frontier methods (i.e., parametric versus non-parametric methods). They found that the central tendencies were similar, with figures around 80%. Overall, the non-parametric methods gave lower efficiency estimates and had greater dispersion of the efficiency values than the parametric methods.

The information comparing the efficiency rankings of DMUs across measurement methods is very limited. Only a few studies applied both parametric and non-parametric methods to the same data (e.g., Drake and Weyman-Jones, 1996; Ferrier and Lovell, 1990; Resti, 1997; Giokas, 1991; Sheldon, 1994 and Sheldon and Haegler, 1993). Only three of these studies reported comparisons of efficiency rankings. The study by Drake and Weyman-Jones (1996) reported that although the efficiency results were quantitatively very different between parametric and non-parametric approaches, (the parametric model suggested that most of the variation in performance was due to random

error), both methods provided extremely similar efficiency rankings. The results from the study by Ferrier and Lovell (1990) were quite the opposite. Both techniques yielded very similar results relating to average efficiency levels, but the correlation between the efficiency rankings was not significantly different from zero. Resti (1997) reported that the results did not differ dramatically between the methods and a high positive correlation between efficiency rankings of individual DMUs was identified. The other studies only reported comparisons between the efficiency results. Giokas (1991) reported that the cost efficiency estimate and its partition between technical and allocative components differed significantly between the methods. Similarly, Sheldon (1994) and Sheldon and Haegler (1993) reported that the efficiency results were rather different between both techniques.

Overall, it seems clear that more research is needed into the effect of the measurement methods on the efficiency estimates and rankings within the banking context. Managerial, policy and research issues may be more convincingly addressed if more than one frontier technique is applied to the same set of data to demonstrate the robustness of the results obtained. Finding an agreement between competing methods would certainly increase the confidence in the results.

3.4.2 Sources of banks inefficiency

Because of the differences in the methodologies employed (e.g., parametric versus non-parametric techniques), the nature of the institutions, and the characteristics of the banking systems (e.g., different countries with different regulatory systems), the results of the studies are often contradictory. Thus, most results cannot be generalised beyond the specific setting in which they were derived and only a few general conclusions can be brought forward.

The main source of inefficiency appears to be of managerial (pure technical) nature, with scale and allocative inefficiencies accounting for only a small proportion of the inefficiencies detected. Berger *et al.* (1993) first noted the conclusion relating to the relative importance of technical inefficiency. They reported that “*technical inefficiencies account for on the order of 20% or more of costs in banking, while scale and product mix inefficiencies, when they can be accurately estimated, are usually found to account for less than 5% of costs*”.

To date, the number of studies that measured allocative efficiency is quite limited and more evidence is needed before the results on allocative efficiency can be generalised with greater confidence.

A limitation of the analysis of scale efficiency and returns to scale is that the empirical studies typically postulate the returns to scale properties of the production technology, without providing some evidence to support the choices made. As a result, the conclusions of studies regarding scale efficiency and returns to scale characteristics may be misleading due to an artificial fitting of a VRS frontier to a data set with an underlying CRS technology (or vice-versa).

For example, a few studies of the US financial system found conflicting results regarding the returns to scale of the institutions: Rangan *et al.* (1988) and Aly *et al.* (1990) concluded that most banks were operating under CRS. Ferrier and Lovell (1990) found that most banks exhibited IRS, whilst Miller and Noulas (1996) found that most banks were in the DRS region. This is still a controversial issue deserving further analysis.

3.4.3 Definition of inputs and outputs and sensitivity of the efficiency measures to variable specification

The difficulty and the controversy that surrounds the definition and measurement of the inputs and outputs of banking, as discussed earlier in this chapter, is reflected in the diversity of the input-output sets used in the empirical studies. These are still very much influenced by the analyst's own view of the banking activity, the overall objectives of the study, and the availability of data. Nevertheless, most studies followed the general guidelines of either the production or intermediation approaches.

This section discusses the insights that empirical studies have brought to the conceptual models for the definition of inputs and outputs discussed earlier in the chapter.

In relation to the output specification under the production approach, some studies compared the efficiency results obtained with outputs measured by numbers of accounts versus the financial values of these accounts. The study by Berg *et al.* (1991) only found small differences in the distribution of efficiency estimates and mean efficiency values due to the alternative output specifications. The study by Kuussaari (1993) also found similar distributions of the efficiency scores, but the mean efficiency was higher when financial values were used as outputs. In addition, both studies found that the ranking of the DMUs was significantly affected by the choice of the output measure. Overall, the results indicate that the efficiency estimates can be importantly affected by how the output is measured.

To overcome these problems, some authors have used more than one measure to enable the full characterisation of the output associated with the business activity. Schaffnit *et al.* (1997) included in the output set simultaneously a measure of the number of transactions processed and the number of accounts handled. In this case, the number of

accounts is used as a proxy for the maintenance activities undertaken by the bank. Others used both the number of accounts and their average size (see Ferrier and Lovell, 1990) or the total balance of accounts and their average size (see Zaim, 1995).

In relation to the on-going discussion about the role of deposits under the intermediation approach, the study by Favero and Papi (1995) has analysed the impact on the efficiency estimates of treating the deposits either as inputs or as outputs. It was found that efficiency was somewhat higher when deposits were specified as outputs. The Spearman rank correlation coefficient between the efficiency rankings obtained under both specifications was rather high (i.e., 0.77). Since the way the deposits are treated in the models can affect the individual DMUs efficiency estimates, this aspect of model specification should be carefully dealt with.

Irrespective of the approach adopted, the output measures specified in the empirical studies generally omit three important aspects of the banking activity: the off-balance sheet business, the quality of customer services, and the risk inherent to bank loans.

With the increase in competition among financial institutions, the banks' intermediation margin is reducing and the importance of off-balance business is growing. The input-output set of the efficiency studies should be broadened accordingly, in order to include the cross-selling activities, the financial resources captured through non-traditional products such as investment funds, securities or insurances, and the commissions obtained from financial services provided to customers. These aspects of banking activity have been included more often in recent empirical studies, such as: Tulkens (1993), Sheldon *et al.* (1993), Sheldon (1994), Drake and Howcroft (1994), Schaffnit *et al.* (1997) and Athanassopoulos (1998).

Only a few studies by Athanassopoulos (1997), Parken (1987) and Soteriou and Stavrinides (1997) analysed the impact of service quality on banking efficiency. With the highly competitive environment currently faced by the banks, service quality is becoming a core element of differentiation required to attract new customers and maintain current ones. Due to its contributing to the long-run viability of banks, quality indicators should be more often included in the analysis of financial institutions. Some of the reasons why these measures are generally left out of the input-output set of efficiency assessments are the difficulties encountered in defining service quality appropriately and finding reliable data to quantify it.

Risk is an important feature associated with bank loans. By excluding it from the data set of efficiency studies, a bank may be considered more efficient just as it increases the volume of loans granted by allowing higher risk levels. Several approaches to the inclusion of risk in the efficiency models have been proposed in the literature. Charnes *et al.* (1990) used provisions for loan losses and actual loan losses as inputs, Brocket *et al.* (1997) used provisions for loan losses as input and allowances for loan losses as output, and Berg *et al.* (1992) used (negative) loan losses as an output to reflect the quality of loan evaluations. Berg *et al.* (1992) compared the efficiency results after the introduction of (negative) loan losses in the output set with the results without the risk-related variable. It was found that the inclusion of the risk indicator in the output set only caused minor changes to the efficiency results.

Although the value of provisions or actual loan losses should be used with a time lead, since the losses arise mostly from loans granted in previous years, the length of the lead is hard to establish, and consequently all studies have used data from the period under

analysis. Future research should attempt to include some *ex-post* measure relating to losses due to bad credits granted during the period under analysis.

3.4.4 Orientation of the assessment

Most banking efficiency studies have adopted an input oriented perspective, showing that the downsizing perspective is predominant among bankers and researchers. The reason generally presented for the choice of this perspective is that financial institutions are considered to operate within a certain market, where they have no direct control over the amount of services their customers require. Therefore, the objective of most assessments is to determine the ideal use of resources for the volume and mix of services currently required.

The output orientation, which focuses on the potential for business expansion while keeping the current resource levels, may be particularly interesting in growing markets. It can inform management about the extent to which the bank (or bank branches) can handle an increased service volume before requiring additional resources.

Research considering simultaneously both input and output perspectives for the efficiency assessment of financial institutions is still in its infancy. The pioneering work came within the parametric stream of the literature. Berger *et al.* (1993) used a profit function, instead of the usual cost function, to obtain efficiency measures for US banks. With the specification of a profit function it was possible to quantify simultaneously inefficiencies in the input and output sides of the banking activity. This study found that output inefficiencies were on average larger than input inefficiencies, which indicated that the literature might be neglecting an important source of banks' inefficiency. The first attempt to analyse this issue in the non-parametric stream of the literature came from Fare *et al.* (1997), which used the notion of the directional distance function.

3.4.5 The effect of deregulation on banking efficiency

It is often stated that deregulation is intended to improve the efficiency and competitive viability of an industry. Several studies have analysed the impact of deregulation on the efficiency and productivity of the financial services sector, but the results obtained have been mixed.

Norwegian banks experienced improved efficiency but small productivity growth after deregulation (Berg *et al.*, 1992). A similar result was obtained with respect to Swiss banks (Sheldon, 1994 and Sheldon and Haegler, 1993). Conversely, Swedish banks experienced a decline in efficiency during the deregulation period (Hartman and Storbeck, 1996). The main feature of the Turkish deregulation was the reduction of efficiency spread among banks (Zaim, 1995). In India, the deregulation has raised the efficiency levels, particularly for foreign owned banks (Bhattacharyya *et al.*, 1997). In contrast, banking efficiency in the US was relatively unchanged by the deregulation of the early 1980s (Elyasiani and Mehdi, 1995). The results of the studies on the effect of deregulation in Spanish banking were conflicting and largely influenced by the output specification. Measuring the outputs by the number of accounts identified productivity decline, despite improvements in efficiency over time (Griffell-Tatje and Lovell, 1996 and 1997). Conversely, measuring outputs by the value of the accounts found productivity growth due to both technological and efficiency improvements (Griffell-Tatje and Lovell, 1997).

Overall, as illustrated by the examples stated above, the effect of deregulation on the efficiency and productivity of the banking sector seems highly dependent on the specific economic environment of each country. This suggests that the conventional wisdom stating that deregulation always improves efficiency and productivity may be incorrect.

3.4.6 Prediction of financial institution failure

An important role of a country's supervision and regulatory policy is the prevention of banks' failure. This protects the depositors and the financial system from being severely disrupted, causing serious problems to the economic health of a country.

Several studies have attempted to use efficiency measures for predicting financial institutions' failure. This literature has been mainly focused on the US banking sector. It was shown that DEA efficiency scores could be used to differentiate the banks that survive from the failing institutions (Hermalin and Wallace, 1994). Furthermore, the differences in performance can be statistically detected long before failure, as shown in Barr *et al.* (1993). Other studies have successfully used the efficiency scores to improve the predictive accuracy of models currently used by regulators (see Barr *et al.*, 1994 and Wheelock and Wilson, 1995). This research is an important contribution to inform government policy regarding financial institutions.

3.4.7 Organisational form and ownership impact on efficiency

Financial institutions can be organised in a number of different ways, according to the country's regulation. The studies of the impact of organisational form on efficiency have focused on US institutions.

In the US banking sector, large organisations can choose between two main types of organisational forms: a multibank holding company, where a commonly owned group of banks have separate charters and financial books, versus an extensive branch banking arrangement, merged into a single charter with a consolidated operation. The impact of organisational form on efficiency has been studied by Grabowski *et al.* (1993). The results suggested that branch banking might lead to greater efficiency levels than keeping banks separate within a multibank holding company.

Smaller US financial institutions can choose between unit and branch organisational forms. The study by Aly *et al.* (1990) did not identify significant efficiency differences between these two types of organisational forms.

Concerning the impact of public versus private ownership on efficiency levels, the empirical studies referring to Belgium (Tulkens, 1993) and Turkey (Zaim, 1995), indicated that public banks are more efficient than private banks in both countries. Chen (1998) obtained the opposite results referring to Taiwanese banks. The study revealed that the efficiency of publicly owned banks is lower than that of the recently created privately owned banks. The main source of inefficiency in the public banks is of scale nature. The study of Indian banks by Bhattacharyya *et al.* (1997) indicated that prior to liberalisation public banks were more efficient than privately owned banks and foreign owned banks. However, this picture changed after liberalisation. The performance of foreign-owned banks improved, privately owned banks maintained their performance levels, whereas publicly owned banks experienced a decline in performance.

3.4.8 Mergers impact on efficiency

In recent years there has been a sharp increase in the number of mergers and acquisitions between financial institutions. This has put the spotlight on the efficiency implications of bank mergers.

The efficiency effect of mergers constitutes a particularly important issue, since merger applicants often cite prospective efficiency benefits as a justification for merger approval. The conventional wisdom, supported by bank consultants and the popular press, suggests that mergers can be successful in reducing costs and improving efficiency. However, more rigorous academic studies do not generally find evidence to support this argument. Most of the frontier analysis studies on bank mergers used

parametric models. The only study using non-parametric models analysed the Spanish financial sector (Grifell-Tatje and Lovell, 1996). It did not find evidence of productivity gains following mergers and acquisitions. Other studies based on parametric models include: Akhavein *et al.* (1997), Altunbas *et al.* (1997), Benston *et al.* (1995), Berger and Humphrey (1992), Fixler and Zieschang (1993), Peristiani (1997), Rhoades (1993 and 1998) and Vennet (1996).

3.4.9 Inter-country comparisons

The banking markets of all countries are becoming increasingly integrated due to the globalisation of financial business enabled by information systems and technological innovations. Inter-country comparisons of financial institutions can provide valuable information regarding the consequences of this progressive integration of the financial systems, as competition is expected to become stronger.

However, the comparison of banking efficiency and productivity across countries was rarely attempted. This may be due to the difficulty in collecting comparable data from several countries. In addition, the cultural, regulatory and economic environments faced by the banks are different, and a few problems can arise when incorporating this type of information into the efficiency models. Clearly, more research is needed in this important topic, to enable a sharper comparison of frontiers corresponding to different countries, taking into account the specificity of the environments faced by the institutions.

To date, only four studies have compared banks across countries using frontier methods (e.g., Berg *et al.*, 1993; Bergendahl, 1998; Pastor *et al.*, 1997; Allen and Rai, 1996). Berg *et al.* (1993) compared the efficiency and productivity of Nordic banks, using two different approaches: a Malmquist-type index and a 'common frontier' with pooled data

from all countries. Bergendahl (1998) also compared the efficiency of Nordic countries, but using only a 'common' frontier approach. Pastor *et al.* (1997) compared the banks in seven European countries and the US using a Malmquist-type index. Allen and Rai (1996) reported an inter-country study that compared efficiency levels for banks in 15 countries using a parametric model.

3.4.10 Improving the performance of branches within a financial institution

This section discusses the use of frontier methods to improve managerial performance. This involves the identification of benchmarking DMUs with high efficiency levels, and encouraging inefficient DMUs to attain a set of targets leading to enhanced performance.

Most banks have their own internal benchmarking procedures, often based on the use of accounting and operational ratios. However, such methods typically lack a powerful and comprehensive optimising technique to identify targets leading to performance improvements. The studies that used frontier methods to complement existing internal benchmarking procedures have shown their usefulness as a tool for informing performance improvement decisions. These studies primarily focused on performance improvements have assessed the relative efficiency of branches within a single institution.

Unlike the studies at the institutional level, which compare banks within a financial market, the studies of bank branches almost exclusively used non-parametric frontier methods. This is justified by the fact that most studies had detailed proprietary data available, such that measurement errors and data problems were not likely to affect significantly the data. The only exception to the use of non-parametric models for the analysis of bank branches' efficiency was the study by Berger *et al.* (1997), which was based on the Distribution Free Approach.

Taking advantage of the physical flow data often available (e.g., number of transactions processed, number of hours worked, etc.), the input-output set is usually defined based on the production approach.

Many studies performed *ex-post* analysis to identify the main determinants of branch efficiency. To date, the results have not been very informative, possibly due to the lack of detailed data regarding the context in which the branches' business takes place. Almost all studies that regressed the efficiency estimates on a set of explanatory variables have only been able to explain a relatively small portion of the total variation. In addition, as noted by Berger and Humphrey (1997), the practice of regressing efficiency values on other explanatory variables may lead to misleading results. Although in the context of financial institution assessments the central tendency of the efficiency values is generally similar across frontier techniques, the rankings of DMUs can differ. Since rankings may differ depending on the frontier technique used, the robustness of the efficiency estimates should be extensively tested before attempting to fit a regression line.

A fundamental limitation of branches' assessments is the inability to distinguish between the inefficiencies that can be attributed to poor management versus those that come with the market. The latter are reflected in customer demographics and economic conditions of the surrounding area, over which management has little control. This issue is particularly important, as the market conditions surrounding a branch can have a significant influence on the activity levels, as shown by Soenen (1974), Clawson (1974), Doyle *et al.* (1981) and Boufounou (1995).

The environmental characteristics outside managerial control were ignored in most efficiency assessments due to the difficulty in collecting adequate data. In order to avoid

the problems caused by measuring relative efficiency of DMUs facing non-homogeneous environments, most of the earlier studies on branch efficiency were based on a relatively small sample. However, as small samples limit the discriminatory power of DEA, some of the latest studies have increased the sample size and included branches operating under different market conditions. Thus, the homogeneity assumption underlying the estimation of relative efficiency has been overlooked in most of the recent studies. Methodological enhancements to account for the effect of the environment on the activity of the branches are in special need. Controlling for this influence will enable a truer picture of what management can do to improve performance.

Another issue particularly relevant for the improvement of financial institution performance is the quantification of the potential impact of efficiency changes on profitability levels. To date, the relation between efficiency and profitability has not yet been analysed in depth. Only the studies by Drake and Howcroft (1994) and Schaffnit *et al.* (1997) analysed the relation between branches' efficiency and profitability, concluding that more efficient branches tend to be more profitable. Miller and Noulas (1996) obtained a similar result at the bank level.

Finally, banks operate in a constantly changing environment, where the products and services provided must evolve in response to competitive pressures and the requirements of increasingly sophisticated customers. Nowadays more and more financial services are provided through computer and telephone, with an increasing number of customers willing to do their banking business through these media. In such a context, the future of bank branches is unclear, as their mix of products and services will have to change in response to these new delivery channels. Thus, it may be necessary to expand branches' assessments to include the newly emerging areas of their operations.

3.5 Summary and conclusions

During the last 15 years, the measurement of banking efficiency has been the focus of much research. The extensive literature on this topic reflects the growing interest on financial institutions' performance. This chapter reviewed the literature, summarising the main aims, methodological issues and conclusions of previous studies on bank's efficiency. The information gathered contributed to the identification of the issues that deserve further attention. Some of these issues were addressed in the empirical part of this thesis by proposing enhanced models and methods for efficiency assessment and improvement.

In the frontier analysis literature, there is a competition between Data Envelopment Analysis and parametric models. In the context of banking business, the parametric model adopted in most empirical studies has been the Stochastic Frontier Approach. Both DEA and SFA have advantages and disadvantages, and the choice of the most appropriate method should depend on the objectives of the analysis and the quality of the data available. The SFA performs well when certain underlying assumptions hold true. However, this approach should be avoided when the characteristics of the technology of production are not known *a-priori* or when they cannot be tested. Due to the minimal assumptions underlying the DEA model, the ability to consider multiple inputs and outputs, and the possibility of decomposing the efficiency estimate into the different sources in a simple manner, DEA is generally accepted as a more appropriate tool for assessments using clean data sets. This justifies the use of the DEA method in this thesis. The major problem associated with the DEA approach is the sensitivity to outliers and random error.

The models and methods developed in this thesis for the assessment and improvement of bank branches' performance were built taking into account the state-of-the-art relating to the issues reviewed in this chapter. The relation between some of these issues and the following chapters is highlighted next.

The *choice of inputs and outputs* for a banking efficiency assessment is a controversial topic. The approaches most often used in empirical studies are the production and intermediation approaches. These two approaches serve different purposes, and should be used together to gain a better understanding of the DMUs' performance whenever there is sufficient data to implement such a research design. Chapter 5 addresses this issue by providing an overall characterisation of bank branches' efficiency using both approaches and discussing their relationship. Nevertheless, each approach has a specific scope, so that the results can be particularly valuable for certain purposes. The last chapters of the thesis are concerned with branches' operational activity, and therefore the input-output set was defined based on the production approach.

It is fair to say that no bank seems to rely exclusively on one approach for internal performance assessments. Ratio analysis, profitability measures and frontier methods are all used in financial institutions. These should be considered as complementary approaches for performance assessment. This thesis illustrates how DEA can complement the profitability measures frequently used in financial organisations. A discussion of the potential *impact of operational and intermediation efficiency improvements on profitability levels* is included in Chapter 5.

The identification of the major *sources of inefficiency* is a perennial topic in banking circles. However, there seems to be no single satisfactory answer, as different institutions have different features. This thesis contributes to this discussion by comparing the main

sources of inefficiency under both the production and intermediation approaches (see Chapter 5). Chapter 6 discusses in greater detail the impact of pure technical and scale efficiency in branches' operational activity, whereas Chapter 7 focuses on cost efficiency issues.

In terms of the characterisation of bank branches' performance, previous studies were often criticised for not being able to *distinguish between the inefficiencies that can be attributed to poor management from those that come with the market*. This issue is addressed in Chapter 8. It develops an enhanced method for the comparison of performance of groups of DMUs operating in different conditions. In terms of bank branches' assessments, this method involves the use of a new index for the comparison of regional performance. This index can be decomposed into an index for the comparison of within-region efficiency spread (reflecting managerial performance) and an index comparing the productivity of the best-practice frontiers (reflecting environmental conditions and policies within which the branches are required to operate).

CHAPTER 4

Introduction to the Portuguese banking sector and the case study of the Portuguese bank

4.1 Introduction

This thesis aims to provide a comprehensive framework for performance measurement and improvement in financial institutions. The enhanced methods and models proposed throughout the thesis were motivated by the analysis of a particular bank, in order to guarantee that the developments proposed are relevant and well adjusted to the needs of today's organisations.

Given the recent evolution of the financial markets in Europe and most industrialised countries towards greater liberalisation, linked to the globalisation of financial business, the critical issues faced by the financial institutions are increasingly homogeneous. Indeed, the organisations are competing in the same global financial market, created by the use of modern information systems.

Under such circumstances, by choosing an institution operating in a liberalised financial market, under the model of universal banking, the framework developed for the assessment and improvement of performance is potentially applicable to other financial institutions in different countries, with only minor adjustments.

The institution chosen for the analysis reported in this thesis is a Portuguese commercial bank whose origins date back from last century. Its history reflects the evolution of the Portuguese banking sector. This bank considers improving efficiency and productivity a

crucial issue to retain competitiveness, and thus it makes a particularly interesting case study.

The purpose of this chapter is to provide the background to the Portuguese financial sector and introduce the bank used in the case study. Particular emphasis is given to the performance measurement methods currently used in the bank. These will be complemented in the later chapters of this thesis by frontier analysis methods, with the objective of obtaining a comprehensive framework, based on optimising methods, that can lead the bank to increased efficiency and productivity levels.

4.2 Overview of the Portuguese banking sector

4.2.1 Introduction

The financial sector is becoming the most significant economic sector in modern societies. In the US, the financial sector accounts for almost 5% of the Gross Domestic Product (GDP) and employs 5.4 million people. The Portuguese financial sector accounts for approximately 7% of Portugal's GDP and 2% of total employment in the economy¹. Similar statistics are found in other European countries, e.g., 5.5% of the GDP in Germany and 3.5% in Italy. In the UK and Switzerland, major international financial centres, the financial sector accounts for 12% and 9% of the countries' GDP, respectively.

In the last twenty years, the Portuguese financial system has undergone substantial changes. It has evolved from a system largely dominated by the state and strongly regulated, to an open, liberalised and competitive market, led by private institutions.

¹ Source: *National Accounts* (1986-1992, 1993, 1994 and 1995) published by the National Statistics Institute, Portugal (in Portuguese).

The transformations of the Portuguese banking sector, though shared by other banking systems in the European Union, were among the most significant, considering the starting point and the pace of introduction of the changes required by the liberalisation process to establish the single European financial market.

Before 1974, during the Portuguese dictatorial period, the banking sector, as most of the economy, was effectively closed to the European and other foreign markets. Almost all banks were privately owned, with only one banking institution owned by the State².

In 1974 and 1975, following the political transformations in Portugal, the main sectors of the economy were nationalised. All banking institutions were nationalised, with the exception of the three foreign banks existing in Portugal at the time³ and other specialised institutions with a mutual character. This resulted in eight nationalised banks, seven of which came from large financial groups established in the 60s and the eighth from the merger in 1975 of three similar banks. At that time, the state-owned banks accounted for virtually the whole market, whilst the few foreign banks had a marginal business share.

The 1975-1983 period was characterised by heavy government intervention all over the economy, which made the Portuguese banking sector one of the most regulated western European markets, with credit ceilings, administratively set interest rates, strict international capital movement controls and legal entry barriers. Competitiveness was substantially limited by the government, with the objective of financing a large budget

² *Caixa Geral de Depósitos* (CGD), which dates back from the nineteenth century (1876) and has always been part of the public sector.

³ *Crédit Franco-Portugais*, *Banco do Brazil* and *Bank of London and South America*.

deficit and maintaining economic stabilisation through dominance of the monetary policy.

All banks were subject to credit ceilings (only abolished in 1991), defined according to their liability base. This policy suppressed competition in the loans market. The credit ceilings forced banks to maintain a large proportion of their assets in idle applications (as excess reserves), which were available to finance the government deficit at a very low cost (Borges, 1993).

In addition, the public banks were often forced to offer credit to unsuccessful nationalised companies, without much prospect of repayment. This created a high level of bad loans in the banks.

Interest rate regulation imposed a minimum rate on term deposits (to encourage savings), a maximum rate on demand deposits (to protect less efficient institutions) and a maximum rate on loans (to prevent excessively high rates, which could be generated by the scarcity of loanable funds resulting from the credit ceilings).

The banking activity was closed to private initiative and new foreign banks could not enter the Portuguese market. Branching was also highly regulated. Until 1986 all new branches depended on the central bank's authorisation. For every four new branches, a bank had to open a fifth in a less-favourable region designated by the Bank of Portugal. This represented a significant cost penalty for the expansion of the branch networks.

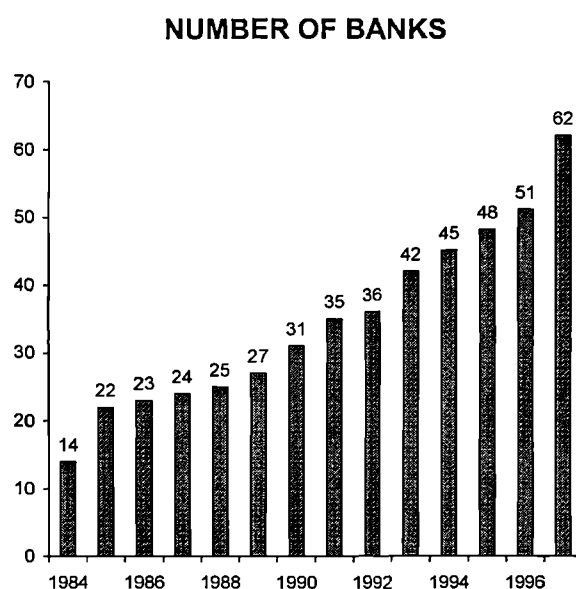
A new period in the evolution of the Portuguese financial sector started in 1984, with the authorisation of new private institutions to operate in the banking sector. The creation of new Portuguese banks and the entry of foreign banks into the Portuguese market were the drivers of change, introducing a competitive spirit in the banking industry.

However, the full impact of competition and deregulation was only felt after the entry of Portugal in the European Community, in 1986. The entry in the European Community made the liberalisation of the financial sector imperative, in line with community wide directives for the harmonisation of regulation in the financial services. The liberalisation process consisted of the lifting of most legal constraints on banking business, covering three main aspects (see Bank of Portugal, 1997 Annual Report):

- *The structure of banking business:* Restrictions to the composition of assets and to banks' activity were lifted. The compulsive investment in public debt securities was abolished. The legally enforced segmentation of credit institutions was gradually removed, culminating in the legislation adopting the model of universal banking, in late 1992.
- *Competitive conditions:* Competition in the banking sector ceased being controlled directly by the state through limits to credit granting and to the establishment of new banks or expansion of branch networks. Also, the administrative regulation of interest rates ended in 1992.
- *Banks international activity:* The liberalisation of the financial system at the domestic level was carried out alongside the gradual liberalisation of capital flows within the European Community, between 1986 and 1992. The full accomplishment of this process resulted in the free establishment and supply of financial services in the European Community area.

4.2.2 Structure of the banking sector

As a result of the liberalisation policies, the structure of the banking sector changed significantly. In 1984 fourteen banks operated in Portugal. In 1985 one financial institution was changed to bank status, which became the first Portuguese private bank after the revolution in 1974. New foreign banks were progressively authorised, but remained small. Between 1984 and 1989, 13 new banks started activity, virtually doubling the number of banks operating in Portugal. From 1989 onwards the number of banks has grown substantially, reaching 62 at the end of 1997, see Figure 4.1. In recent years, most of the newcomers in the business do not supply retail services. Instead, they operate in the stock market and in the investment banking segment.



*Figure 4.1 – Number of banks operating in Portugal between 1984 and 1997
(source: Bank of Portugal, 1997 Annual Report)*

After the opening of the banking activity to new private and foreign institutions in 1984, the number of bank branches increased significantly, not only due to the creation of new retail banks but also due to the expansion of branch networks by the “old” institutions. The restrictions to the growth of the bank branch networks have gradually decreased since the beginning of the decade and all restrictions were finally abolished in December 1992.

In 4 years, between 1988 and 1992, the number of bank branches almost doubled. After that, due to the stiffer competition for the provision of financial services, the number of branches operating in Portugal continued to increase steadily, at an average rate of 8% per year (see Figure 4.2). These branches are highly concentrated in the four major Portuguese cities (i.e., 58% of the total number of branches).

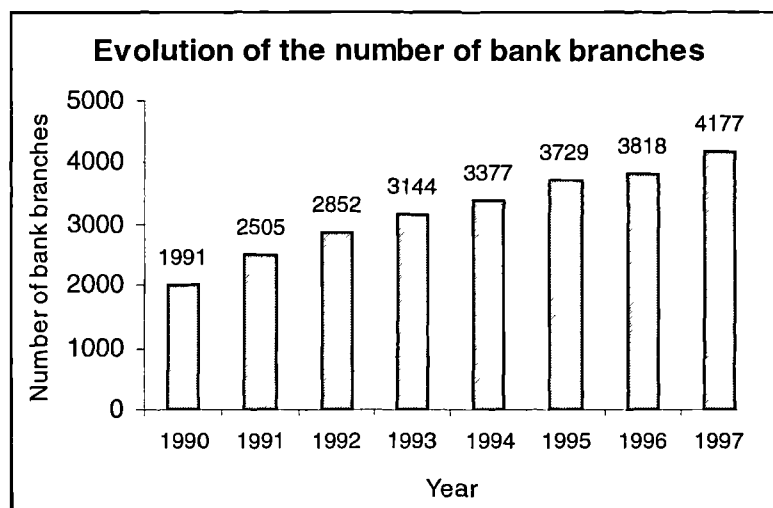


Figure 4.2 – Evolution of the number of bank branches operating in Portugal (source: Official report of the Portuguese Banking Association, June 1997 and Bank of Portugal, 1997 Annual Report).

This expansion of the networks occurred alongside a virtual stabilisation to the number of employees, which resulted in a reduction to the average number of employees per branch, e.g., from 15.7 in 1996 to 14.4 in 1997. For the year 1996, the population served by each branch was, on average, 2600 people. At the European level, this number is around 3850 people⁴, which suggests that the Portuguese banking sector may be occurring in “overbranching”, (see *Official report of the Portuguese Banking Association, June 1997* and *Bank of Portugal, 1997 Annual Report*).

Supported by technological innovation, banks have enhanced the production and distribution of financial services, which has resulted in the gradual substitutions of labour for new technologies, e.g., telephone banking, ATMs, and other forms of self-service or remote supply of services. This resulted in lower staff costs, i.e., the ratio of staff costs to total operating expenses fell from 76.4% in 1985, to 69.2% in 1991 and 63.1% in 1997. However, the operating costs have kept on growing in recent years (9.1% growth in 1996

⁴ Germany and Austria have the highest values of population served (i.e., 10900 people per branch) and Belgium has the lowest value (i.e., 1300 people per branch).

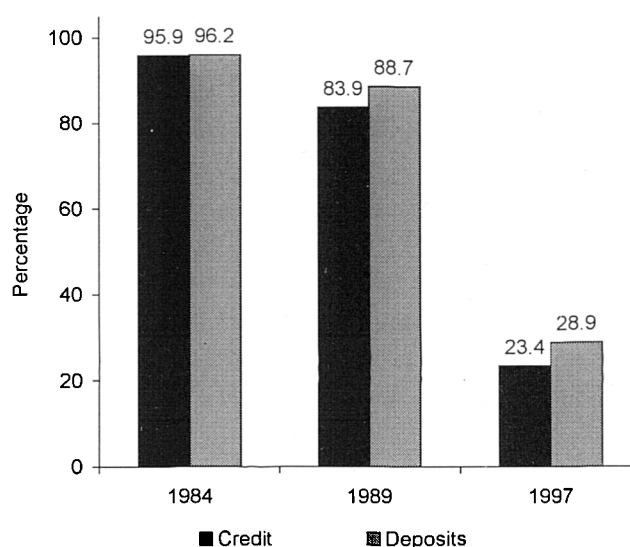
and 1997)⁵. This is due to the strong investment in information and communication systems, increasingly considered a decisive factor for the development of banks' competitiveness.

4.2.3 Market share of the institutions

In 1984, the banks owned by the State accounted for more than 95% of the market. By the beginning of 1989 the State continued to hold the bulk of banking business and was still the owner of banks representing almost 90% of the market.

The privatisation of the state-owned banks started in 1989. Their market share fell to about 45% in 1993 and stood below 30% at the end of 1997, see Figure 4.3.

MARKET SHARE OF STATE-OWNED BANKS



*Figure 4.3 – Market share of state owned banks
(source: Bank of Portugal, 1997 Annual Report)*

The sequence of mergers and acquisitions within the banking sector, intensified since 1994, has changed the structure of the banking sector towards greater concentration (see Figure 4.4). These mergers and acquisitions were often justified by the banks involved as

⁵ Source: *Bank of Portugal, 1997 Annual Report*.

a means to cut costs and increase efficiency through the sharing of operating structures, managerial capabilities and taking advantage of potential economies of scale.

MARKET SHARE OF THE FIVE MAJOR BANKING GROUPS

	Assets	Funds from clients	Credit to clients	Net income for the year
1993.....	67.7	72.4	71.9	79.4
1994.....	68.8	71.8	69.7	87.8
1995.....	77.6	80.6	76.0	90.3
1996.....	83.0	84.9	79.4	93.1
1997.....	78.0	82.8	77.1	85.4

Note:

(a) Sum of the market shares of institutions belonging to the five leading banking groups, in each of the variables considered and on a non-consolidated basis.

*Figure 4.4 – Market share of the five major banking groups
(source: Bank of Portugal, 1997 Annual Report)*

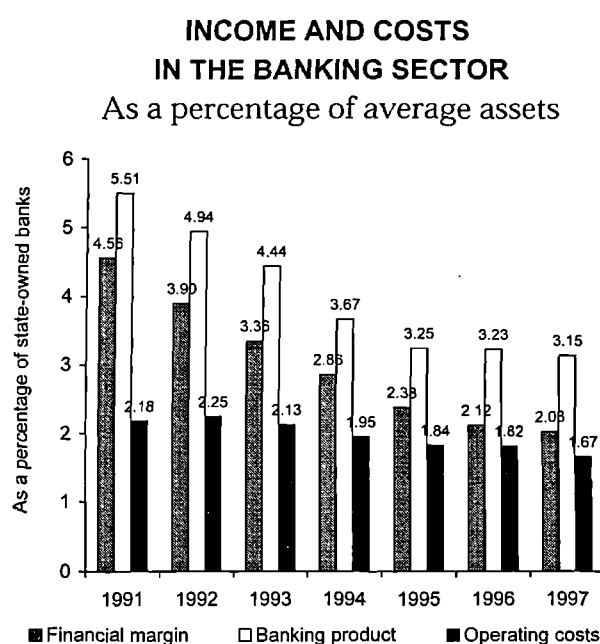
In 1993, the market share of the five major Portuguese banking groups was 67.7% (in terms of assets). This value increased to 83% in 1996⁶. This illustrates the effort of the Portuguese banks to achieve a scale size that enables them to compete with larger financial institutions within the European Union area and to survive in the increasingly globalised financial market.

⁶ In 1996, the ranking (in terms of assets) of the five major banking groups was the following: the group *Caixa Geral de Depósitos* had the first position. The group *Banco Comercial Português / Banco Português do Atlântico* occupied the second position. The group *Banco Pinto e Sotto Mayor / Banco Totta e Açores* had the third position in the ranking in terms of assets, the second position in terms of funds from clients and credit to clients and occupied a lower position in terms of net income. The group *Banco Espírito Santo* occupied the fourth position, although it had a higher position in terms of net income. The group *Banco Português de Investimento* had the fifth position in the ranking, with a market share of approximately 10%. Of particular relevance during 1996 was the acquisition of *Banco de Fomento e Exterior* and *Banco Borges e Irmão* by the *Banco Português de Investimento*. (For further details see the *Official Report of the Portuguese Banking Association*, June 1997).

After the significant rises in business concentration observed in 1995 and 1996, the year 1997 was characterised by a reduction in the market share of the major banks. This resulted from the growth above the sector average of some medium-size retail banks. It should be noted that the degree of concentration of the Portuguese banking sector is still below the values of other countries of similar size. Thus, further mergers and acquisitions are likely to occur in the near future.

4.2.4 Competitive conditions and financial margins

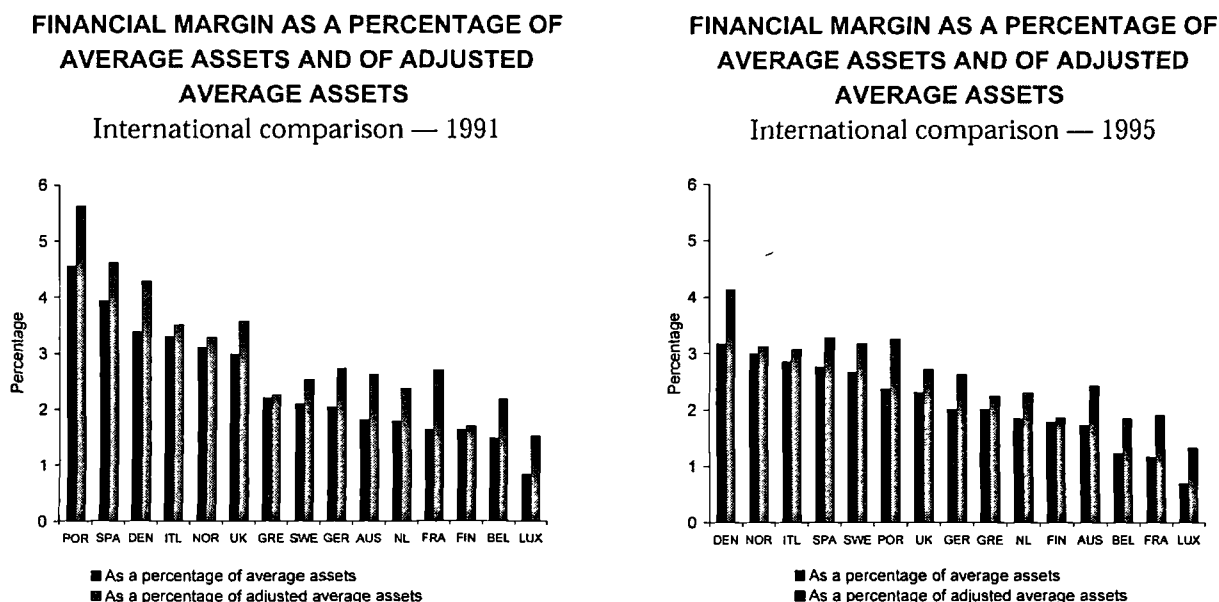
The liberalisation of the interest rates (completed in May 1992) and the elimination of the credit ceilings⁷, contributed to an increase in competition in the sector, where the prices practised in banking operations assumed strategic importance. As a result, the intermediation (or financial) margin decreased substantially, from 4.6% of average assets in 1991 to 2.0% in 1997, see Figure 4.5.



*Figure 4.5 - Income and costs in the Portuguese banking sector
(source: Bank of Portugal, 1997 Annual Report)*

⁷ Substituted in 1991 by a liquidity control regime based upon market mechanisms.

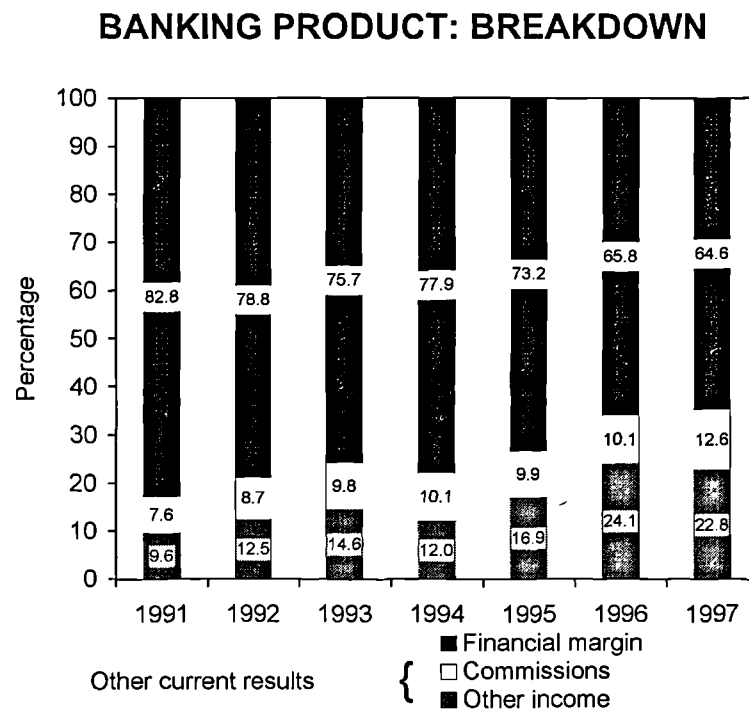
The evolution of the financial margin in Portugal between 1991 and 1997, (expressed as a percentage of average assets), placed this indicator at similar levels to those of most countries in the European Economic Area, see Figure 4.6.



*Figure 4.6 – International comparison of financial margins
(source: OECD, bank profitability, Paris 1997)*

In 1991 Portugal had the highest value of the financial margin. In 1995, the latest year for which data is available for a wide set of countries, the Portuguese financial margin was within the average values for the European countries.

In parallel to these changes, the structure of banks' profits also changed significantly, with the financial margin (i.e., profit generated from the traditional intermediation activity) decreasing its importance against the profits from other activities (i.e., commissions and results from other financial operations). In 1991, the share of commissions in the banking product was 7.6%, whereas it reached 12.6% in 1997 (see Figure 4.7). Despite this increase, the current value is still below those observed in most European countries.



*Figure 4.7 – Breakdown of the banking product
(source: Bank of Portugal, 1997 Annual Report)*

4.3 The case study: introduction to the Portuguese bank

This section introduces the case study of the Portuguese bank that was used as a vehicle for developing performance measurement and improvement methods for banking institutions. It was considered that the development of enhanced models and methods motivated by the analysis of a particular bank would result in a contribution best adjusted to the reality faced by the institutions operating in today's financial markets. The models and methods proposed throughout the thesis are illustrated using the data from this case study, giving a better understanding of their strengths and weaknesses as decision support tools for the management of bank branch networks.

A Portuguese commercial bank whose origins date back from last century was chosen for the case study. This bank was nationalised in 1974 and reprivatized in 1991. It was acquired by a financial group created in 1981, which has since become the fourth largest in the Portuguese banking sector. Its current size was attained through organic growth as

well as mergers and acquisitions. Its history reflects the evolution of the Portuguese banking sector, and in this context we believe it makes a particularly interesting case study. Background information concerning this bank is detailed in the remainder of this chapter.

4.3.1 Description of the bank

The case study used in this thesis refers to a bank founded in 1861. It has been engaged in several mergers since its creation, which contributed to gain a good reputation both as a retail bank with a strong link to the industry and as an institution with talent to perform operations in the financial market, with a large experience with Government securities.

In 1974 the bank was nationalised, following the political changes that had occurred in Portugal. This resulted in a deterioration of the bank's economic situation. In 1986 it started a process to regain economic and financial soundness, with a view towards reprivatisation. The reprivatisation occurred in 1991, when the sixth largest Portuguese financial group at that time bought the bank. This resulted in a substantial strengthening of the equities, which enabled a strategy of growth and modernisation to be followed.

The branch network has seen a substantial growth since 1991 (Figure 4.8), alongside gains in productivity (see Figure 4.9 and Table 4.1). The expansion of the network involved a redesign of branches' structure, towards smaller scale size and greater support from centralised activities. The new branches were opened with 4 employees, on average. Most of these employees (82%) resulted from staff reallocations within the bank.

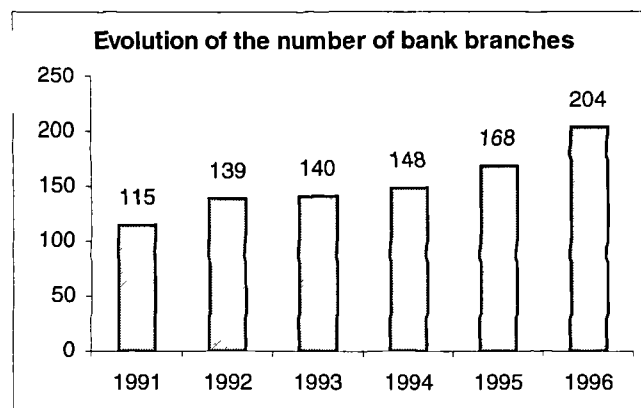


Figure 4.8 – Branch network expansion

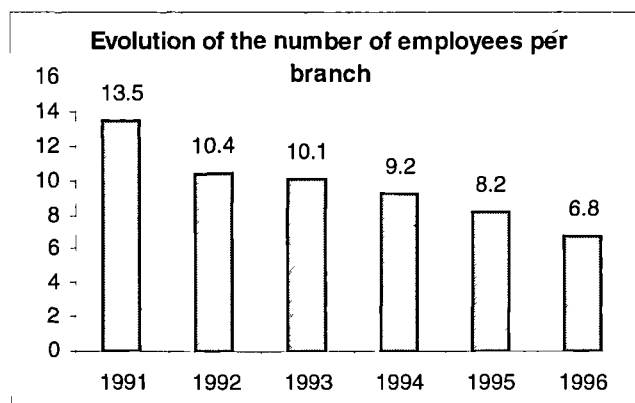


Figure 4.9 – Evolution of the number of employees per branch

Table 4.1 – Evolution of branch employees' productivity

<i>Branch employees' productivity</i> (monetary values in million escudos)	1993	1994	1995	1996
<u>Total deposits</u> No. employees	410.9	478.2	519.2	562.8
<u>Total credit</u> No. employees	58.5	91.3	126.7	143.9

The analysis of the branch network reported in this thesis refers to data from the year of 1996. The original sample included 168 branches, corresponding to those in activity at the end of 1995.

In terms of the commercial activity for the year of 1996, the bank has reported the following strategic directions⁸:

- *To intensify the effort of diversification of sources of revenue. In particular, to increase the value of revenues not related to the financial margin.*
- *To increase productive efficiency, taking into account the best practices of internal and external competitors. This objective involves an increase in revenues, as well as the implementation of actions leading to cost reductions.*
- *To attain a scale size that enables maintaining a competitive position among the main institutions of the Portuguese financial sector.*

The first objective is particularly important given the substantial reduction in financial margins observed in the Portuguese banking sector and the suppression of exchange business within the European monetary union area. In 1996, the application of customer funds as off balance sheet resources (e.g., investment funds, securities or insurances of capitalisation) represented a third of the total value of the resources from clients. In recent years the relative value of commissions in the banking product has increased substantially for most banking institutions. This trend is expected to continue in coming years.

The second objective of increased efficiency is seen as an essential requirement to maintain a strong competitive position.

The achievement of the third objective involved the acquisition of the second largest financial group owned by the State, whose privatisation occurred during 1996. This operation was seen as a means to gain competitiveness and take advantage of economies of scale.

⁸ Stated in the 'Report and Accounts' of the bank for the year of 1996.

The analysis that is reported in this thesis is a contribution to the achievement of these objectives.

During 1998, another bank was acquired by the group, which complemented its distribution channels giving access to the “in-store banking” business. All the banks of this financial group were merged during 1998.

As this thesis focuses on performance assessment and improvement of bank branch networks, the next section describes the performance measures used in the bank at the time of the study reported in this thesis.

4.3.2 Performance measurement methods at the Bank

The bank used two different methods to assess the branch network performance. One was based on the evaluation of profitability at the branch level, and called *Earning Analysis System* (EAS). The bank used the EAS for the first time during 1996. The second method is called *System of Incentives and Motivation* (SIM), and has been in use since 1993.

Before describing both methods in greater detail, the managerial structure of the branch network is introduced next, for the year under analysis (1996). As illustrated in Figure 4.10, its organisation is mainly geographical.

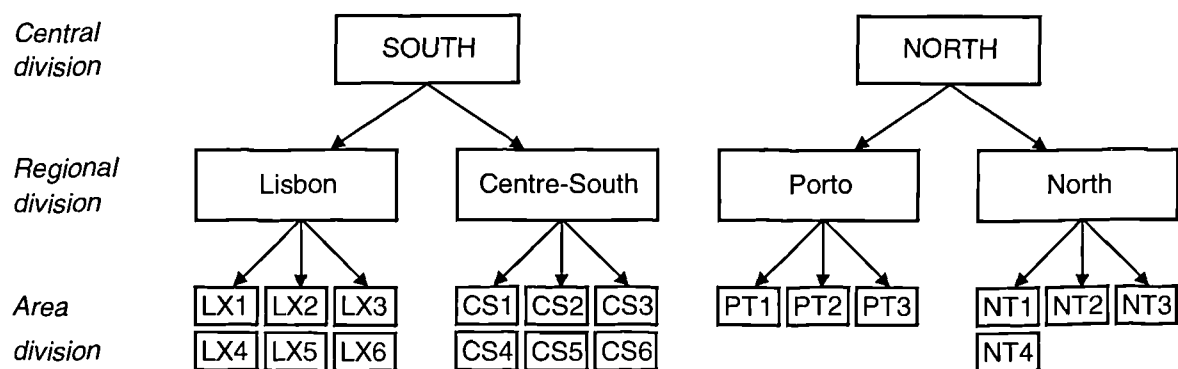


Figure 4.10 – Managerial structure of the branch network

These branches only deal with individuals and small business accounts and their activities are reasonably homogeneous. They are scattered across the country, although the two main Portuguese cities have a higher concentration of branches. The network is separated in two central managerial divisions (e.g., North and South). These are further divided in regional and area divisions. Each area division has a manager responsible for the operational activity of a group of branches (that varies between 7 and 10). The branches in the same area division are usually located close to each other.

4.3.2.1 Earning Analysis System

The Earning Analysis System (EAS) derives the monthly and annual profitability of the bank. The profitability measure is obtained separately for each of the bank's business units (e.g., credit cards unit, corporate business unit, retail business unit, etc.). In particular, the analysis of the retail business unit includes the analysis of profitability at the branch level.

The profitability measure has three main components: *intermediation margin*, *second margin* and operational costs. Its derivation is described next.

The calculation of the intermediation margin requires an estimation of the volume of loans and deposits at each branch, as well as their corresponding financial margin. The financial margin (or spread) for deposits is obtained as the difference between the interest rate received from the application of the money in the interbank monetary market and the interest rate paid to customers for their deposits. The financial margin for loans is obtained as the difference between the interest rate received from branch customers and the interest rate paid to the interbank monetary market for the use of the funds.

The second margin refers to revenues beyond the traditional intermediation activity. It is based on net revenues from off-balance sheet business (such as commissions for cross-selling insurances, securities or investment funds) and other type of income from financial business.

The sum of the intermediation and second margins gives the Banking Product of the branch.

The operational costs include labour costs and costs of services and supplies by external companies, adjusted for the costs of interbranch transactions. The purpose of the adjustment is to offset net expenses incurred in servicing customers of other branches. For example, downtown branches often cash checks for customers whose accounts are credited to the suburban branches where the accounts were opened. If reckoning of these interbranch transactions were not made, downtown branches would appear to incur higher costs per volume of accounts than suburban branches. The cost of interbranch transactions is calculated as a net value that is positive if the branch performs fewer transactions for others than it receives, and negative otherwise.

The branches' cash-flow of exploration is obtained as the sum of the intermediation and second margins, deducted from operational costs. This is the measure of profitability used in this thesis, although the bank refines it further by including amortisations, variation in provisions, indirect costs and extraordinary results. However, as these components are not directly related to the branches' operational activity, we have chosen to exclude them from the performance assessment described in this thesis.

4.3.2.2 *System of Incentives and Motivation*

The System of Incentives and Motivation (SIM) focuses on the branches' volume of business. Every four months, the bank's senior management sets global targets for the branch network. These are specified for each of the products and services considered of strategic importance. The global targets are then distributed equitably by the bank branches. In general, branches with higher values of accounts and larger number of clients are expected to contribute more to the attainment of the overall objectives. The best branches in the achievement of the objectives set receive a bonus.

The SIM objectives are set for a 4-month period and include several items (e.g., volume of current and saving deposits, investment funds sold, volume of credits, cross-selling of insurances, etc.). The degree of achievement of each objective is evaluated in percentage terms. This percentage is then multiplied by the weights attached to each objective, so that each branch receives a given number of points. The first and second classified (i.e., the branches with the highest number of points) in each regional division (i.e., Lisbon, Centre-South, Porto and North) are considered the benchmarks and receive a bonus.

4.3.3 Motivation to use the DEA method

The EAS and SIM methods are important tools to analyse branches' performance, which address different purposes. The SIM is a valuable tool for establishing a clear incentive scheme for branch staff. The EAS provides the bank with a clearer picture of the key costs and revenues of branches' activities, as well as the business profile of each branch. However, both the SIM and EAS have limitations as methods for performance assessment.

The SIM targets do not take into account directly the costs of delivering services to customers. Also, as the targets are based on past performance, the branches that have

been poor in the past may be favoured and set less demanding targets. Most importantly, setting targets for sales motivation and bonus is quite different to planning a branch network for efficient operation.

In relation to the EAS, profitability measures alone ignore the branches' efficiency and potential for improved performance. Also, due to their accounting nature, they are difficult to relate to the operational activity and be accepted by staff.

This thesis proposes the use of DEA alongside the EAS profitability measure to provide an overall picture of branches' performance and identify directions for improvement. The resulting information can complement the SIM method for setting business volume targets, as the DEA measure has the advantage of simultaneously taking into account the volume of business and the corresponding level of resource consumption.

4.4 Summary and conclusions

This thesis is focused on the development of a framework for performance measurement and improvement in financial organisations. It is based on the analysis of a case study of a commercial bank, with the objective of ensuring that the methods and models developed are relevant and well adjusted to the needs of financial institutions.

The purpose of this chapter was the characterisation of the Portuguese financial sector, where the bank used for the motivation and illustration of the developments proposed in this thesis operates. The chapter also introduced the bank, giving particular emphasis to the methods currently used for the analysis of branch performance.

For more than a decade, Portuguese banks operated under a highly regulated market, where the State controlled most of the institutions, interest rates for deposits and loans were set by the authorities, entry in the market was banned, and opening of new branches

depended on the central bank's authorisation. This conjuncture led to a situation of virtually no competition in the market, which is likely to have resulted in low efficiency levels at the Portuguese financial institutions.

A progressive deregulation process started in 1985, and today the financial sector is undergoing strong competitive pressures. In this context, the ability of the banks to adapt to the new situation is crucial to retain viability. However, so far the number of studies on the efficiency of the Portuguese financial sector is very limited, and none has analysed efficiency at the organisational level. Given the evolution of the financial sector, we believe that the study of the efficiency of a Portuguese bank makes a particularly interesting case study, as increasing efficiency levels is seen as an essential requirement to maintain a strong competitive position and a decisive factor to ensure economic viability.

CHAPTER 5

Efficiency and profitability measurement in bank branch networks

5.1 Introduction

This chapter develops an overall framework for the assessment of performance of financial institutions. It addresses a set of issues related to the management of bank branch networks.

The first issue concerns the enhancement of the DEA models to simultaneously account for input and output inefficiencies with a cost minimisation perspective. To date, the cost minimisation models used in banking efficiency assessments were based on the efficiency concept of Farrell (1957). These models search for the input levels that could minimise costs given current input prices and output levels. This chapter proposes the use of an alternative model, also with a cost minimisation perspective, whose difference in relation to previous approaches is to enable changes to the DMUs' output levels. This model provides a measure of the efficiency in the generation of the current level of total revenue with minimal cost, given the input and output prices and the characteristics of the production technology. This approach is based on the Fare, Grosskopf and Lovell (1994) model for *revenue indirect input cost efficiency* measurement.

The model used in this chapter searches for the minimal cost for each branch that still enables generating its current total revenue level. No increase to total revenue is allowed because there is no evidence that the market could support any expansion to business

levels. Also, from a short-term perspective taking business from competitors is difficult, as customers generally have a loyal relation with their main bank. However, the output mix of each DMU is considered to some extent a result of managerial efforts, as the customers may be persuaded to take a different mix of financial products, e.g., investing their savings in securities rather than keeping term deposits.

The importance of considering both input and output sides of the DMUs' activity in the assessment of economic efficiency was emphasised in Berger *et al.* (1993). The parametric model¹ used for the assessment of financial institutions found that the output inefficiencies were on average larger than the input inefficiencies. This result suggested that the literature was not giving due attention to the output side of banking business.

The research by Fare *et al.* (1997) also highlighted the importance of considering both input and output aspects of banking activity when assessing economic efficiency. Fare *et al.* (1997) used a model based on a directional distance function to assess efficiency with a profit maximisation objective.

These two studies motivated the use of an enhanced DEA model for economic efficiency measurement. The research described here departs from these earlier studies of financial institutions because no revenue enhancements are allowed when searching for profit maximisation. All profitability gains must be achieved through the rationalisation of costs as, in the case study analysed, the market potential of each DMU was considered fully explored.

¹ Based on the specification of a profit function.

The second issue analysed in this chapter relates to the choice of the inputs and outputs for the assessment of bank branches' efficiency. This chapter compares two approaches based on different concepts of bank branches' activity: the 'production' and 'intermediation' approaches. By using the two approaches in the analysis of a bank branch network, this study contributes to the clarification of the issues that can be addressed with each of these approaches, and to illustrate how they can jointly be used to inform managers with respect to branches' performance.

Although in recent years bank branches' efficiency has attracted widespread attention, only a few studies (Athanasopoulos, 1997; Oral and Yolalan, 1990; Oral *et al.*, 1992) explored branches' efficiency under both the 'production' and 'intermediation' perspectives. The studies by Oral and Yolalan (1990) and Oral *et al.* (1992) found a positive association between the efficiency results obtained in the two approaches, i.e., the branches found efficient in the operational activity, (as modelled by the production approach), also performed well with respect to financial business, (as modelled by the intermediation approach²). The study by Athanasopoulos (1997) focused on target setting and service quality issues, and it did not compare the results obtained in the two approaches.

The third issue addressed in this chapter concerns the efficiency implications of bank mergers in terms of the branch networks. One way in which improvements could occur is by a superior management team taking over the branch network and running it more efficiently. The ability to rationalise costs through this process depends in large part on whether banks can centrally control costs at the individual branches. If local management

² In the papers by Oral and Yolalan (1990) and Oral *et al.* (1992), the authors called this approach a 'profitability assessment'.

has a significant impact on branches' efficiency, then the success of the implementation of new policies and procedures defined centrally can be limited by the quality of local management.

The alternative channel for mergers to improve branches' efficiency is through consolidations in which one branch is closed and its business is transferred to another branch. Efficiency gains may be expected if the business of an inefficient branch can be transferred to a nearby branch that is more efficiently managed. However, the benefits of this type of consolidation can only occur if the merged networks have geographically proximate branches, well below efficient scale size, and with the right conditions available to absorb the closing branches' business.

The analysis described in this chapter is the first to use the DEA method for evaluating the impact of mergers on the efficiency of a branch network. The study by Grifell-Tatjé and Lovell (1996), also based on the DEA method, only explored the impact of mergers on efficiency at the bank level.

The fourth issue addressed in this chapter concerns the integration of efficiency and profitability measures in an overall framework for assessing performance at bank branches. It is shown that the efficiency measures are an important complement of profitability evaluations. The joint use of these two indicators can lead to a clearer and more objective performance analysis. In addition, the association of DEA results to profitability measures creates an effective method for directing performance improvements.

Dyson *et al.* (1990) and Boussofiene *et al.* (1991) proposed the use of an efficiency-profitability matrix for the analysis of DMUs' performance. This chapter describes an empirical application of this method to the assessment of a bank branch network.

The structure of this chapter is as follows. Section 5.2 describes the framework for the efficiency assessment and the extensions to the DEA method that enable a simultaneous evaluation of input and output efficiencies within a cost minimisation framework. Section 5.3 discusses the branches' activity models, the choice of the inputs and outputs for the efficiency assessments, and the data used. Section 5.4 discusses the empirical findings related to branches' efficiency. Section 5.5 discusses the managerial implications of the empirical results. Section 5.6 integrates the efficiency and profitability measures in an overall framework for performance assessment. Section 5.7 summarises and concludes.

5.2 Framework for the efficiency assessment

This section provides the framework for a comprehensive efficiency assessment of a bank branch network, considering both technical and economic aspects.

In efficiency measurement, the focus can be placed in assessing how well a DMU is using its resources (measured in physical or cost terms) to obtain the current outcome. This corresponds to an input oriented assessment. Alternatively, the focus can be the assessment of how well the outputs (measured in physical or revenue terms) are being produced with the current resources, corresponding to an output orientated analysis. If the optimal is defined in terms of physical production possibilities, the resulting efficiency measure is called technical. Conversely, if the optimal is defined incorporating the economic goals of the DMU, then efficiency is called economic. The most demanding economic objective corresponds to profit maximisation, which may involve both cost reductions and revenue enhancements.

In the context of banking activity, the business potential of a bank branch is very much dependent on the socio-economic conditions of the surrounding area, which is beyond

the decision-makers' control, and persuading customers to change banks is difficult. From a short-term perspective therefore, the business potential can be considered exogenously fixed and thus the efficiency assessment of a bank branch should be done primarily with an input orientation, taking into account the actual level of services required by the customers³.

Nevertheless, even in the short-term, some aspects of branch output may be at the decision-makers' discretion. For example, an account manager may suggest to clients that they invest their savings in securities instead of keeping deposit accounts. As a result, the output mix may be influenced by the commercial strategies followed by branch staff.

This section extends the DEA method further to assess efficiency with an input orientation, enabling trade-offs among outputs provided that the total revenue of each DMU does not change. In the model developed, total revenue is used as a proxy to the DMUs' business potential. This model should be used when the objective of the efficiency assessment is cost minimisation and changes to both input and output mixes are allowed.

The linear programming model associated with the computation of this efficiency measure, referred to as cost-effectiveness in the remainder of this thesis, is described in detail next.

³ This assumption of fixed business levels will be relaxed in a later chapter, when considering branches' performance with a long-term perspective.

Consider a set of n DMUs, $j = 1, \dots, n$, each consuming m inputs, x_{ij} , $i = 1, \dots, m$, to produce s outputs, y_{rj} , $r = 1, \dots, s$. The cost-effectiveness measure can be obtained from a DEA 'weights' model using input and output weights restrictions based on the observed input and output prices at each DMU. This model is shown in (5.1).

Cost-effectiveness model: 'weights' formulation	(5.1)
$\begin{aligned} \text{Max } \phi_{j_0} &= \sum_{r=1}^s u_r y_{rj_0} \\ \text{s.t. } \sum_{i=1}^m v_i x_{ij_0} &= 1 \\ \sum_{r=1}^s u_r y_{rj} - \sum_{i=1}^m v_i x_{ij} &\leq 0, \quad j = 1, \dots, n \\ v_{i^a} - \frac{p_{i^a j_0}}{p_{i^b j_0}} v_{i^b} &= 0, \quad i^a < i^b, \quad i^a, i^b = 1, \dots, m \\ u_{r^a} - \frac{\rho_{r^a j_0}}{\rho_{r^b j_0}} u_{r^b} &= 0, \quad r^a < r^b, \quad r^a, r^b = 1, \dots, s \end{aligned}$	
(5.1a)	

In the above model $p_{i^a j_0}$ and $\rho_{r^a j_0}$ represent the input (i^a) and output (r^a) prices at DMU j_0 , respectively. The value of $\phi_{j_0}^*$, at the optimal solution to (5.1), gives the measure of cost-effectiveness for DMU j_0 .

Comparing the 'weights' formulation of cost-efficiency (model (2.23) in Chapter 2) and cost-effectiveness (model (5.1)), the latter includes an additional restriction (5.1a) related to the output weights. As a result, whilst cost efficiency searches for the input levels that enable the production of *current outputs* at minimal cost, cost-effectiveness identifies the input and output levels that enable the generation of *current total revenue* at minimal cost. Since the output weights used in the assessment must reflect the output prices at each DMU (as imposed by (5.1a)), the cost-effectiveness assessment allows trade-offs

between the outputs provided they generate a similar level of total revenue as the original output combination.

Alternatively, the measure of cost-effectiveness can be obtained from an ‘envelopment’-type formulation, based in the model proposed by Fare, Grosskopf and Lovell (1994, p.146) for the measurement of Revenue Indirect Input Cost Efficiency. The model presented in (5.2) calculates the minimal cost leading to the attainment of current revenue levels, given the input and output prices at DMU j_0 ($p_{i^0 j_0}$ and $\rho_{r^0 j_0}$, respectively) and the underlying characteristics of the technology of production.

The model in Fare, Grosskopf and Lovell (1994) differs from the model below as it computes the minimal cost of obtaining a pre-specified target revenue. If the pre-specified target revenue used in the Fare, Grosskopf and Lovell (1994) model is set to be equal to the current revenue generated by the DMU j_0 under assessment, all models (i.e., model (5.1), model (5.2) and Fare, Grosskopf and Lovell (1994) revenue indirect input cost efficiency model) would compute a similar efficiency measure.

Model to determine the minimum cost associated with cost-effectiveness		(5.2)
$\text{Min } \psi_{j_0} = \sum_{i=1}^m p_{ij_0} x_i^0$		
Subject to		
$\sum_{j=1}^n x_{ij} \lambda_j = x_i^0,$	$i = 1, \dots, m$	
$\sum_{j=1}^n y_{rj} \lambda_j = y_r^0,$	$r = 1, \dots, s$	(5.2a)
$\sum_{r=1}^s \rho_{rj_0} y_r^0 = \sum_{r=1}^s \rho_{rj_0} y_{rj_0}$		(5.2b)
$\lambda_j \geq 0, x_i^0 \geq 0, y_r^0 \geq 0, \quad j = 1, \dots, n; i = 1, \dots, m; r = 1, \dots, s$		

In the above formulation, the value of $\psi_{j_0}^*$, at the optimal solution, gives the minimal cost value that enables the attainment of the current level of total revenue at DMU j_0 .

x_i^{0*} and y_r^{0*} are variables that, at the optimal solution to (5.2), give the amount of input i and output r that would make DMU j_0 cost-effective.

This model does not allow any increase to the current level of total revenue because there is no evidence that the business potential of the DMUs' surrounding area could enable such revenue enhancements. This is consistent with the assumption of fixed business levels stated previously. In addition, the model does not allow revenue reductions either, as this could imply profitability reductions, which are certainly not desirable.

The relation between the models for the identification of the minimal cost associated with cost-effectiveness (model (5.2)) and cost efficiency (model (2.20) in chapter 2) is

that the output restriction of the latter (e.g., $\sum_{j=1}^n y_{rj} \lambda_j \geq y_{rj_0}$, which only allows target output levels greater or equal to the current output levels) is relaxed to allow further changes to output levels. In model (5.2), the output restriction above is replaced by two restrictions, (5.2a) and (5.2b), such that the target for some outputs may be less than the current value, provided that the total revenue of DMU j_0 remains the same⁴ (i.e., equal to

$$\sum_{r=1}^s \rho_{rj_0} y_{rj_0}).$$

⁴ Note that both in the cost-efficiency and cost-effectiveness models, the input and output targets must correspond to a feasible position in the production possibility set, as imposed by the first two restrictions of these models.

Based on the use of model (5.2), the measure of cost-effectiveness is obtained as the ratio of the minimum cost obtained at the optimal solution to (5.2) to the current cost at DMU j_0 , as follows:

$$\text{Cost effectiveness } j_0 = \frac{\psi_{j_0}^*}{\sum_{i=1}^m p_{ij_0} x_{ij_0}} \quad (5.3)$$

The cost-effectiveness measure can be decomposed into cost efficiency and output mix efficiency. These measures can be defined more precisely as follows (note that all measures assume fixed input and output prices):

- Cost-effectiveness measures the ability of a DMU to achieve current revenue levels at minimal cost, allowing for changes to input-output levels and mix.
- Output mix efficiency measures the extent to which a DMU produces the outputs with the right mix to enable the attainment of the current level of total revenue with minimal cost.

Output mix efficiency is obtained as the ratio of the minimal cost of DMU j_0 ($\psi_{j_0}^*$), obtained at the optimal solution to model (5.2), where output levels can be changed, to the minimal cost of DMU j_0 obtained at the optimal solution to model (2.20) in Chapter 2, where the output levels are fixed.

- Cost efficiency measures the ability to produce current outputs at minimal cost, allowing for changes to input levels and mix.

The relation between cost-effectiveness, output mix efficiency and cost efficiency is as follows:

$$\text{Cost - effectiveness} = \text{Output mix efficiency} \times \text{Cost efficiency} \quad (5.4)$$

Finally, cost efficiency can be multiplicatively decomposed into pure technical efficiency, scale efficiency and input allocative efficiency⁵. Note that scale efficiency is only defined for DMUs operating under VRS.

- Pure technical efficiency measures the amount by which all inputs can be proportionally reduced without reducing the outputs (allowing for variable returns to scale).
- Scale efficiency measures the ability to adopt the most productive scale size.
- Input allocative efficiency measures the extent to which the DMUs use the right input mix to produce current outputs at minimal cost.

In the remainder of this chapter, these efficiency measures will be used for a comprehensive characterisation of the efficiency status of the bank branch network. The relation between the five measures defined above is summarised in Figure 5.1.

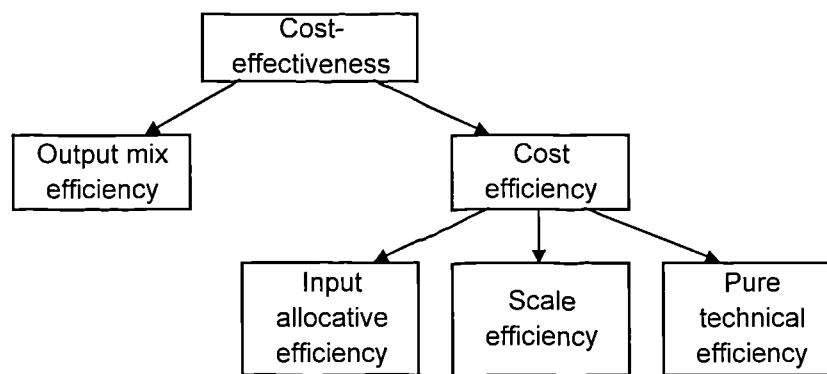


Figure 5.1 – *Decomposition of the cost-effectiveness measure*

5.3 Branches' activity models

One of the crucial issues to build a model for the assessment of bank branches' efficiency is the identification of appropriate inputs and outputs. This issue is not straightforward and an extended and unresolved controversy remains in the literature. The two approaches most often used are the production and intermediation approaches.

⁵ These efficiency measures were described in greater detail in Chapter 2.

Most of the previous studies on bank branches' efficiency chose an input-output set based on the production approach. This shows that the focus of the efficiency analysis has often been the commercial/operational aspects of branches' activity. To date, only a limited number of studies reported efficiency assessments based on the intermediation approach (e.g., Athanassopoulos, 1997; Oral and Yolalan, 1990; Oral *et al.*, 1992; and Soteriou and Zenios, 1999). This suggests that the branches' financial performance is more often assessed with traditional techniques, such as financial ratios, than with modern techniques based on frontier methods.

This chapter aims to provide a comprehensive assessment of bank branches' efficiency. It shows that the joint use of the production and intermediation approaches can provide a complete characterisation of branches' performance. The two main aspects of branches' activity, relating to the ability to provide transactions/document-processing services to customers and the ability to serve as financial intermediaries between savers and investors at a profit, are best captured by a research design based on both approaches. Nevertheless, each approach has a specific scope, and the results obtained with any of these approaches are particularly useful for certain purposes.

The following sections describe the inputs and outputs used for modelling branches' activity based on the production and intermediation approaches. These were defined in collaboration with bank management, to ensure an accurate representation of branches' activity.

The empirical results were obtained from the analysis of branches' activity during 1996. Only those branches with full information for all variables of the production and intermediation approaches were analysed, yielding a final data set of 144 branches. This data is likely to be considerably cleaner than standard banking data sets, as it was

collected from branches of a single bank, based on information from an internal accounting system designed to monitor branch profitability.

5.3.1 Production approach

The inputs and outputs defined for the production approach were as follows:

- Inputs :**
- Number of branch and account managers.
 - Number of administrative and commercial staff.
 - Number of tellers.
 - Operational costs (excluding staff costs).
- Outputs:**
- Total value of deposits.
 - Total value of loans.
 - Total value of off balance sheet business.
 - Number of general service transactions.

In relation to the **input** set, staff is the key resource used. In the bank branch network used as a case study, personnel expenses represented about 75% of total branches' operating expenses. The employees were separated in three categories according to their functions at the branch (i.e., branch/account managers, administrative/commercial staff and tellers). The fourth input considered refers to operational costs, excluding staff costs (e.g., supplies and services from other companies, commissions paid and other operational costs).

The **output** set included four factors: total value of deposits, total value of loans, total value of off-balance sheet business negotiated at the branch⁶, and the number of general service transactions performed during the period under analysis. Account values were preferred to account numbers as a proxy of the maintenance activities and other specialised services provided to customers. The general service transactions included in

⁶ E.g., sales of investment funds, insurances of capitalisation, or other types of cross-selling activities.

the output set consisted of twenty different types of transactions, with a similar level of complexity, all performed mostly by tellers and administrative/commercial staff⁷. All transactions involving error corrections or transactions done by automated means, without involving human intervention, were excluded from the output measure.

Table 5.1 provides summary information on the input and output sets and input prices (i.e., salaries) used in the production approach. The prices were obtained for each individual branch. The salary corresponds to the annual value, including fringe benefits. All monetary variables are measured in million escudos.

The output prices (i.e., spread of financial products and net commissions) are also reported in Table 5.1 to test the ability of the model to provide an accurate picture of branches' performance. This is done by testing whether the variables used in the production approach can provide a good estimate of the branches' overall results (e.g., cash-flow) when associated with other variables complementing the operational perspective by reflecting the financial aspects of the business activity.

⁷ (1) Drafts at credit, (2) drafts at debit, (3) purchase of foreign currency cheques, (4) cheques cashed, (5) collection of taxes, (6) collection of rates, (7) discounts, (8) deposits in cash, (9) payment of expenses, (10) devolution, (11) delivery of values, (12) purchase of foreign bank notes, (13) sale of foreign bank notes, (14) other credits, (15) other debits (16) withdrawals of savings accounts, (17) payment of remunerations, (18) purchase of petrol vouchers, (19) money transfers at credit, (20) money transfers at debit.

Table 5.1 - Summary information of variables related to the production approach

<i>Variables</i>	<i>Symbol</i>	<i>Mean</i>	<i>Standard deviation</i>
Inputs			
No. branch/account managers	X _{1P}	3.5	1.4
No. administrative/commercial staff	X _{2P}	3.3	2.0
No. tellers	X _{3P}	2.3	1.0
Operational costs	X _{4P}	18.1	10.1
Outputs			
Total value of deposits	Y _{1P}	3662.1	3042.5
Total value of loans	Y _{2P}	1214.8	1020.5
Total value off balance sheet business	Y _{3P}	1347.8	1268.7
No. general service transactions	Y _{4P}	92421.6	42509.2
Input prices			
Salary of branch/account managers	PX _{1P}	5.38	0.57
Salary of administrative/commercial staff	PX _{2P}	3.37	0.47
Salary of tellers	PX _{3P}	4.22	0.39
Output prices			
Spread of deposits	PY _{1P}	2.19%	0.54%
Spread of loans	PY _{2P}	5.04%	0.90%
Commissions from off-balance sheet business (%)	PY _{3P}	2.90%	2.34%
$\text{Cash-flow} = Y_{1P} \times PY_{1P} + Y_{2P} \times PY_{2P} + Y_{3P} \times PY_{3P} - (X_{1P} \times PX_{1P} + X_{2P} \times PX_{2P} + X_{3P} \times PX_{3P} + X_{4P})$			

Table 5.1 shows how the cash-flow of the branch can be estimated from the data included in the production approach and additional information on output prices. The correlation between the actual cash flow of the branch (obtained by the EAS – *Earning Analysis System*) and the estimate based on the variables used in the production approach, complemented with data relating to the results of the financial intermediation, is very high: 0.989. This indicates that the input-output set defined for the production approach is a good representation of branches' operational activity.

The production approach focuses on the evaluation of branches' operational activity and staff performance. This approach should be used when analysing branches' operational and commercial efficiency.

Conversely, the intermediation approach is more appropriate for evaluating branches' economic viability, as it captures the essence of bank branches' activity as financial

intermediaries. This approach is more inclusive, as it considers the total costs incurred (i.e., operational and interest costs) for a certain level of business. Therefore, it should be used when analysing strategic issues such as closing branches or defining their target business mix.

The input-output set defined for the analysis of branches' activity based on the intermediation approach is described in the next section.

5.3.2 Intermediation approach

The inputs and outputs specified for the intermediation approach were as follows:

- Inputs :**
 - Non-interest costs.
 - Interest costs from deposits.
 - Interest costs from loans.
- Output set A:**
 - Total value of deposits.
 - Total value of loans.
 - Total value of off balance sheet business.
- Output set B:**
 - Total revenue from deposits.
 - Total revenue from loans.
 - Commissions from off balance sheet business.

For the intermediation approach, three **inputs** were specified, i.e., non-interest costs, interest costs from deposits and interest costs from loans. The *interest costs from deposits* correspond to the interest paid to customers as a compensation for the use of their funds. The *interest costs from loans* correspond to the interest paid to the interbank monetary market for borrowing the funds made available to customers through loans. The *non-interest costs* include the commissions paid and operational costs, adjusted for the costs of interbranch transactions. The purpose of this adjustment is to offset branches' net expenses incurred in servicing customers of other branches and the bank as a whole. For example, downtown offices often cash checks for customers whose

accounts are credited to the suburban branches where the accounts were opened. If reckoning of these interbranch transactions were not made, downtown offices would appear to be inefficient for incurring more costs per unit of their own accounts' value than suburban offices. The adjustment consists in adding to the non-interest costs the net costs of interbranch transactions. This net cost is positive if a branch performs fewer transactions for others than it receives and negative if the converse is true. With this adjustment, the branches' workload due to general service transactions is accurately accounted for without having to include explicitly the number of transactions in the output set. Therefore, the efficiency measure gives an indication of the ability to minimise costs for the amount of money intermediated (with output set A) or revenue generated (with output set B).

In relation to the outputs, two alternative output sets were specified. Output set A is based on the *monetary value* of branches' business, including the funds intermediated and off-balance sheet resources. Conversely, the output set B is based on the *revenues* generated by branches' business.

In relation to the **output set A**, it includes the *total value of deposits*, *total value of loans* and *total value of off-balance sheet business* of each branch. The value of loans excludes those corresponding to bad credits. This avoids classifying as highly efficient those branches that boosted their credits at the expense of higher risk.

In relation to the **output set B**, the *interest revenues from deposits* correspond to the revenues obtained from applying the money in the interbank monetary market. The *interest revenues from loans* correspond to the interest paid by customers for the use of the funds. The third output corresponds to the *commissions received from the off balance sheet business* of the branch.

In the branch network used as a case study, the revenues earned per unit of the monetary value intermediated, as well as the commissions received per unit of off-balance sheet resources, can vary between the branches. This is because although the interest rates and commissions charged are defined centrally at the bank, the products sold at each branch are different, depending on the customers' requirements. In addition, the branches can negotiate small adjustments to the interest rates and commissions for their core customers. Under such circumstances, the output set B gives a better characterisation of branches' overall efficiency, including both commercial aspects (related to the monetary values intermediated) and financial aspects (corresponding to the revenues earned for a certain volume of business). Conversely, the output set A can be more informative for defining targets relating to the volume of business of each branch. This chapter explores the impact of the different output specifications (e.g., monetary value versus revenue) on the efficiency estimates.

Table 5.2 provides summary information of the variables used in the intermediation approach. All monetary variables are measured in million escudos. The output prices were obtained for each individual branch.

Table 5.2 - Summary information of variables used in the intermediation approach

<i>Variables</i>	<i>Symbol</i>	<i>Mean</i>	<i>Standard deviation</i>
Inputs			
Non-interest costs	X _{1I}	61.6	23.3
Interest costs from deposits	X _{2I}	206.2	183.6
Interest costs from loans	X _{3I}	92.4	77.0
Output set A			
Total value of deposits	Y _{1I}	3662.1	3042.5
Total value of loans	Y _{2I}	1214.8	1020.5
Total value of off balance sheet business	Y _{3I}	1347.8	1268.7
Output set B			
Total revenue from deposits	Y _{4I}	281.2	235.0
Total revenue from loans	Y _{5I}	149.0	115.3
Total value of commissions from off-balance sheet business	Y _{6I}	29.2	20.6
Prices of output set A			
Fund transfer price of deposits (%)	PY _{1I}	7.63%	0.09%
Interest earned from loans (%)	PY _{2I}	12.65%	0.95%
Income from off-balance sheet business (%)	PY _{3I}	2.90%	2.34%
Intermediation model with output set A:			
$\text{Cash-flow} = (Y_{1I} \times PY_{1I} + Y_{2I} \times PY_{2I} + Y_{3I} \times PY_{3I}) - (X_{1I} + X_{2I} + X_{3I})$			
Intermediation model with output set B:			
$\text{Cash-flow} = (Y_{4I} + Y_{5I} + Y_{6I}) - (X_{1I} + X_{2I} + X_{3I})$			

Clearly, the representation of branches' business using output set B is more comprehensive than using output set A, as the latter requires additional information on output prices to enable retrieving the cash-flow generated by the branch.

The correlation between the actual cash-flow of the branch (obtained by the EAS) and the estimate based on the data included in the intermediation approach is also very high: 0.991 (both for output sets A and B). This indicates that the intermediation approach models provide a good representation of branches' activity.

The attractiveness of branches' location, which depends on competition and socio-economic conditions of the surrounding area, was not included in the input sets of the production or intermediation approaches due to data unavailability. However, some input may be necessary for attracting customers in difficult locations and therefore the analysis

may underestimate the efficiency of branches located in such areas. In future analysis with environmental data available, these factors may be incorporated into the efficiency model as uncontrollable inputs. Alternatively, they may be used at a second stage as control variables for the analysis of the impact of location on branches' efficiency.

5.4 Empirical results of the efficiency assessment

5.4.1 Selection of appropriate DEA models regarding returns to scale

It was *a-priori* unclear if branches operational and intermediation activities experienced constant or variable returns to scale. If a restrictive assumption of variable returns to scale is imposed on the models unnecessarily, the resulting efficiency estimates will be greater than the true efficiency values, as the VRS frontier will always provide a closer envelopment of the data. Therefore, it is important to analyse the returns to scale properties of branches' activity prior to the efficiency assessment.

Banker (1996) proposed statistical tests for identifying the type of returns to scale underlying the DMUs' activity. These tests are based on the statistical foundation for the DEA efficiency estimates developed in Banker (1993).

The statistical tests in Banker (1996) analyse the differences in the distributions of inefficiencies derived from the CRS and VRS models. Depending on the assumptions made regarding the true efficiency distribution, different statistical tests can be used. The Kolmogorov-Smirnov two-sample test was chosen for the analysis reported in this section, as due to its non-parametric nature no assumptions need to be made regarding the true inefficiency distribution. See Siegel and Castellan (1988) for further details regarding this type of test.

The *null hypothesis* of the Kolmogorov-Smirnov test states that the distributions of inefficiency obtained using the CRS and VRS models are identical. This implies that the CRS and VRS frontiers are very close and thus scale inefficiency is almost non-existent. This can be interpreted as evidence that the DMUs' activity exhibits CRS.

The corresponding *test statistic* requires the calculation of inefficiencies under the CRS and VRS models for all DMUs. It reflects the differences between the cumulative distributions of the CRS and VRS inefficiency estimates.

If the inefficiency estimates obtained using the CRS and VRS models have a similar distribution, then the differences computed for the test statistic are small. Consequently, there is not enough evidence to reject the null hypothesis and it can be concluded that the DMUs are likely to operate under CRS. Conversely, if the inefficiency distributions are different, the null hypothesis can be rejected and it is concluded that VRS is more likely to exist.

In relation to the results of the Kolmogorov-Smirnov test applied to the bank branches under analysis, the null hypothesis was rejected for the production approach, at a 5% significance level⁸. This indicated that branches operational activity exhibits VRS.

As can be observed in Figure 5.2, the differences between the efficiency distributions obtained from the CRS and VRS models are quite substantial.

⁸ The level of significance of the test statistic obtained (i.e., the p-value) was 0.000.

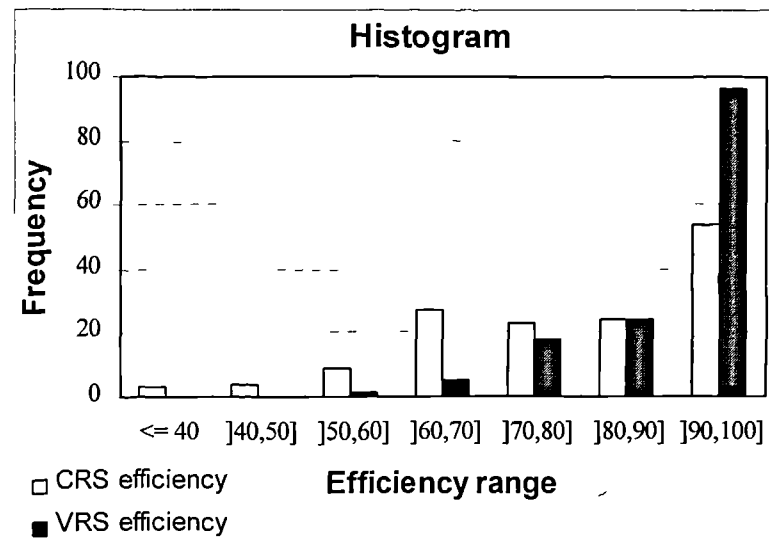


Figure 5.2 – *Efficiency distributions for the production approach under CRS and VRS*

However, for the intermediation approach, there was no evidence to support the rejection of the null hypothesis at a 5% significance level, irrespectively of the output definition used⁹. Thus, it was concluded that the intermediation activity is not significantly affected by branches' scale size.

As shown in Figure 5.3 and Figure 5.4, the efficiency distributions under CRS and VRS are very similar both for the intermediation model with the output set based on business volumes and for the output set based on revenues.

⁹ The level of significance of the test statistic obtained (i.e., the p-value) was 0.600 when using the intermediation approach with the output set based on business volume and 0.051 when using the output set based on revenues.

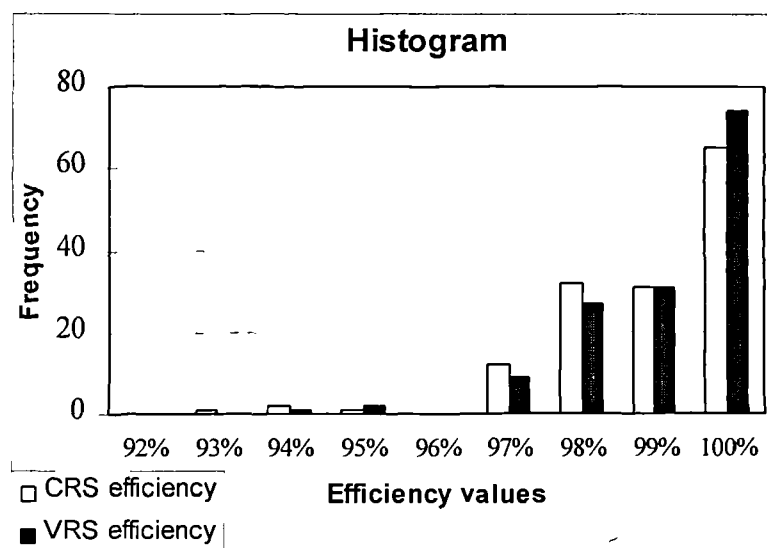


Figure 5.3 - Efficiency distributions for the intermediation approach (output set based on volume of business) under CRS and VRS

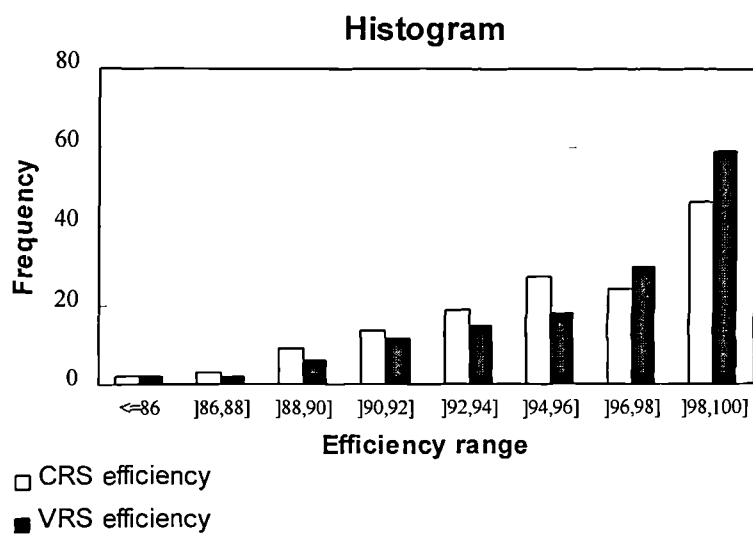


Figure 5.4 - Efficiency distributions for the intermediation approach (output set based on revenues) under CRS and VRS

In conclusion, for the bank branches under analysis, the frontier of the production possibility set should be estimated assuming VRS for the production approach, and assuming CRS for the intermediation approach.

A careful characterisation of the DMUs' returns to scale properties, such as the one described in this section, is essential to obtain robust estimates of efficiency.

The next sections discuss the empirical results regarding branches' efficiency. The results of the production and intermediation approaches are first analysed separately and then compared to discuss their managerial implications. The first measures presented for each approach are aggregate measures, followed by the analysis of their components.

5.4.2 Intermediation efficiency analysis

5.4.2.1 Cost-effectiveness measure

The most comprehensive efficiency measure computed for the intermediation approach was the cost-effectiveness measure. It explores cost reductions through adjustments to input-output levels and mixes. The DEA model used for the calculation of cost-effectiveness was described in detail in section 5.2 (see models (5.1) and (5.2)).

Recall that the branches' short-term business level was considered outside the decision-makers' control. Therefore the target output levels obtained from the cost-effectiveness assessment do not generate a total revenue level exceeding the current level. These output targets can suggest a reorientation of the output mix towards the sales of the banking products that enable the largest cost savings.

These changes to the output mix are only desirable if they enable reductions to total costs (i.e., operational and interest costs). Therefore, the measure of cost-effectiveness was only computed for the intermediation approach.

Furthermore, it is important to analyse the output targets in terms of the best mix of banking products sold at each branch. Thus, the output set of the intermediation approach based on the volume of branches' business is more informative for exploring the managerial targets.

The results of the cost-effectiveness assessment are shown in Table 5.3. These are reported as summary measures for the branch network, using the output set based on business volume.

Table 5.3 – *Cost-effectiveness results for the intermediation approach*

	<i>Cost-Effectiveness</i>
Mean	68.9%
Standard deviation	10.6%
No. of efficient branches	1

The average cost-effectiveness measure is close to 70%. This indicates that, on average, the branches can reduce their total cost to approximately 70% of the current value through adjustments to input-output levels, without reducing the current level of total revenue generated by each branch.

Only one branch is fully cost-effective. This branch can be considered the benchmark of the network in terms of overall efficiency. Note that although having a single benchmark in an evaluation of technical efficiency would be highly unlikely, as this would indicate that the efficient frontier is defined by a single DMU, this results is natural in a cost-effectiveness evaluation. Recall that the inclusion of weights restrictions in the DEA cost-effectiveness model generates an evaluation of efficiency against a ‘value frontier’, which reduces the value of the efficiency estimates and improves the discrimination between the DMUs¹⁰.

The components of the cost-effectiveness measure are described in the next sections.

¹⁰ For example, from an intermediation efficiency perspective, twenty branches from the network analysed were technical efficient, whereas only five were cost efficient and only one cost-effective.

5.4.2.2 Technical efficiency

The results regarding the technical efficiency measure are shown in Table 5.4. They are reported as summary measures for all branches in the sample, using the output set based on business volume.

Table 5.4 - *Results of the technical efficiency measure for the intermediation approach*

<i>Technical efficiency</i>	
Mean	98.5%
Standard deviation	1.3%
No. of efficient branches	20

The average technical efficiency is very high (98.5%). This value indicates there is little scope for equiproportional input reductions, as the branches are already operating close to the frontier of the PPS. Thus, technical inefficiency is almost non-existent in this branch network.

5.4.2.3 Input allocative efficiency

The results of the input allocative efficiency of branches' intermediation activity are shown in Table 5.5, using the output set based on business volume. These are reported as summary measures for the branch network.

Table 5.5 - *Input allocative efficiency results for the intermediation approach*

<i>Input allocative efficiency</i>	
Mean	85.8%
Standard deviation	10.4%
No. of efficient branches	5

Table 5.5 shows that branches' allocative efficiency is on average 86%. This value represents the proportional reduction to the cost associated with a technical efficient point on the frontier of the PPS that can be achieved by adopting the right input mix in light of current prices. This value of 86% shows that there is significant scope for

efficiency improvement by exploring trade-offs between the interest and non-interest costs.

5.4.2.4 Output mix efficiency

The summary results of the output mix efficiency component are shown in Table 5.6. Recall that output mix efficiency evaluates branches' ability to produce the outputs with the right mix to enables the attainment of current revenue at minimal cost. The results reported correspond to the use of the output set based on business volume.

Table 5.6 – *Output mix efficiency results for the intermediation approach*

	<i>Output mix efficiency</i>
Mean	81.7%
Standard deviation	8.8%
No. of efficient branches	1

The average output mix efficiency observed in these branches is approximately 82%. This value represents the average proportional reduction in total branch costs¹¹ that can be attained by adjusting the output mix. It can be concluded that the reorientation of branches' business can bring significant cost cuts to the network.

5.4.2.5 Comparison of efficiency results with different output sets

The results of the intermediation approach reported previously were based on the output set with business volume. This section compares the results of the two alternative output definitions discussed earlier in this chapter, corresponding to *business volume* versus *revenues*.

¹¹ Beyond the cost reductions achieved with the elimination on the inefficiencies detected in the input side (i.e., technical inefficiency and allocative inefficiency).

Table 5.7 shows the average efficiency measures for all branches analysed, for the two alternative output specifications.

Table 5.7 - *Summary of the intermediation efficiency results for the alternative output specifications*

	<i>Output set based on business volume</i>	<i>Output set based on revenues</i>
Cost-effectiveness	69%	---
Output mix efficiency	82%	---
Cost efficiency	85%	87%
Technical efficiency	99%	96%
Input allocative efficiency	86%	91%

Recall that the *cost-effectiveness* of a branch is equal to the product of *output mix efficiency* and *cost efficiency*. Similarly, *cost efficiency* can be obtained as the product of its components corresponding to *technical efficiency* and *input allocative efficiency*.

The analysis of Table 5.7 indicates that the average cost efficiency is only marginally affected by the choice of output measure (e.g., business volume or revenues). The underlying distributions of cost efficiency were also compared using the Kolmogorov-Smirnov test. Although the average efficiency is similar, it was concluded, at a 5% significance level, that the underlying efficiency distributions are different (the p-value of the statistical test was 2.6%).

In relation to the comparison of the efficiency rankings obtained with the alternative output specifications, the Spearman rank correlation coefficient between the cost efficiency measures is rather high (0.8619). This indicates that the efficiency ranking is not much affected by the output specification used.

In relation to allocative efficiency and technical efficiency, the differences between the output specifications are more significant. The differences between the average efficiency values are larger and the results of the Kolmogorov-Smirnov test indicated

that the underlying efficiency distributions have statistically significant differences (at a 5% significance level).

In terms of the efficiency rankings, although the allocative efficiency ranking is not significantly affected by the output measure (the Spearman rank correlation coefficient is 0.773), the technical efficiency ranking is very sensitive to the output specification (the Spearman rank correlation coefficient is only 0.0267).

Overall, it can be concluded that the alternative output specifications do not give dramatically different results in terms of cost efficiency. However, the decomposition of cost efficiency differs substantially between the two approaches. Some reasons for the differences in the results obtained are discussed next.

Based on the output definition corresponding to the volume of business, all branches appear almost fully technical efficient. All the inefficiencies detected were attributed to allocative aspects, suggesting that performance improvements require exploring the trade-offs between the inputs in light of existing prices.

On the other hand, the output definition based on revenues includes implicitly two aspects of banking activity, as the revenues are a function of the amount of money intermediated as well as the prices charged. As a result, some inefficiencies identified as allocative in the first output specification are already reflected in the technical efficiency measure when using the approach based on revenues. This highlights the importance of interpreting the results of the efficiency components in the light of the output specification used.

In conclusion, the choice of the output specification should depend on the purpose of the analysis. It is also possible to choose a research design based on the analysis of both

output specifications. Under such circumstances, if a branch appears more technical inefficient when using an output set based on revenues than with the output set based on business volume, it indicates that the inefficiencies are mostly due to the financial earnings associated with the volume of money intermediated. Conversely, if the branch appears more technical inefficient based on business volume than based on revenues, it may be an indication that the branch concentrates on fewer customers, which reduces the volume of business, although the revenues these customers generate compensate that option. This is a scenario likely to happen in branches located in affluent areas.

5.4.3 Production efficiency analysis

5.4.3.1 Cost efficiency measure

We start the analysis of the empirical results for the production approach with the cost efficiency measure (see Table 5.8). The results are reported as summary measures for all branches analysed.

Table 5.8 – *Cost efficiency results for the production approach*

	<i>Cost efficiency</i>
Mean	69.0%
Standard deviation	18.4%
No. of efficient branches	12

The average cost efficiency of branches' operational activity is 69%. This indicates that with current input prices, the operational costs could be reduced, on average, to 69% of the current values if branches' were fully efficient in the usage of resources.

Similar to previous empirical studies based on the production approach (see the studies by Berger *et al.* (1997) on a US bank, Tulkens (1993) on a Belgium bank, and Athanassopoulos (1998) on a UK bank, that found an average level of CE below 75%), these results also indicate the presence of considerable cost inefficiencies for this

network. Clearly, cost efficiency is not the sole driver in the management of bank branches' operational activity, which can explain the presence of such large inefficiencies.

In order to identify the sources of operational inefficiency, the cost efficiency components (e.g., pure technical, scale and allocative efficiency) are discussed next.

5.4.3.2 *Pure technical efficiency*

Table 5.9 shows the summary results of pure technical efficiency for the production approach. As this efficiency measure accounts for the existence of VRS, the inefficiencies detected are attributable to poor managerial performance.

Table 5.9 – *Pure technical efficiency for the production approach*

<i>Pure Technical Efficiency</i>	
Mean	92.4%
Standard deviation	10.7%
No. of efficient branches	81

The average pure technical efficiency of the operational activity is above 90%. This value represents the equiproportional reductions to current input usage that would enable the achievement of managerial efficiency.

The standard deviation of the pure technical efficiency measure is quite high. The value of 10.7% standard deviation clearly indicates that the efficiency estimates are not dominated by a few outliers that are either very efficient or inefficient. This dispersion of efficiency levels suggests that bank's central management is not able to control tightly branches' operational activity through the definition of procedures or supervision policies. Rather, the quality of local management appears to be quite important in determining branches' operational performance.

5.4.3.3 Scale efficiency

The next step of the analysis was the measurement of scale efficiency. Table 5.10 shows summary information on several scale efficiency indicators for the production approach.

Table 5.10 – *Scale efficiency and related measures for the production approach*

	<i>Scale related measures</i>
Mean scale efficiency	87.9%
Standard deviation	15.8%
No. of scale efficient branches	41
No. (and %) of branches with IRS	88 (61%)
No. (and %) of branches with CRS	41 (29%)
No. (and %) of branches with DRS	15 (10%)
Efficient scale factor:	
Branches with IRS	1.34
Branches with DRS	0.70
Decomposition of scale efficiency:	
IRS efficiency	88.6%
DRS efficiency	99.4%

The data shows that the average scale efficiency of branches is 88%. This value represents the average proportional reduction to the pure technical efficient input levels that could be attained if maximal productivity was achieved.

Table 5.10 also indicates the number of branches with increasing, constant and decreasing returns to scale. It was found that most branches have IRS, i.e., they are below the most productive scale size (MPSS) for their product mixes.

The value of the *efficient scale factor* is also shown. This measure is equal to the ratio of a branch's scale size corresponding to maximum productivity to its actual scale size (see Berger *et al.*, 1997). The actual scale size of a branch is defined in this thesis by the current number of employees working at the branch. The most productive scale size

(MPSS) is defined by the target number of employees after the elimination of both pure technical and scale inefficiencies¹².

For the branches analysed, the average efficient scale factor for branches with IRS is 1.34 and for branches with DRS it is 0.70. This indicates that, on average, branches below MPSS should increase their scale size by 34% to reach scale efficiency and branches above MPSS should decrease their scale size by 30%.

The last measure reported corresponds to the decomposition of the scale efficiency measure of each branch into two parts, corresponding to the scale efficiency associated with IRS (*IRS efficiency*) and the scale efficiency associated with DRS (*DRS efficiency*). The IRS efficiency equals scale efficiency if a DMU is below efficient scale, and equals one otherwise. The DRS efficiency equals scale efficiency if the DMU is above efficient scale, and equals one otherwise (see Berger *et al.*, 1997). Consequently, the product of these efficiency measures (i.e., IRS efficiency and DRS efficiency) equals the scale efficiency measure.

The average IRS efficiency of these branches is 88.6% and the average DRS efficiency is 99.4%. This indicates that most of the scale inefficiency detected is due to increasing returns to scale.

In summary, being at the wrong scale size is a factor that affects substantially branches' operational activity. It can be concluded that the ability to service customer accounts and perform transactions can be done more productively if a certain 'critical mass' of business is achieved, for which the number of transactions and customer account values dealt with by each employee can be increased. The quantification of scale efficiency

¹² The definition of branches' scale size is discussed in more detail in Chapter 6.

showed that most branches are below efficient scale, and typically about 3/4 of the MPSS. Also, the majority of scale inefficiency is attributable to branches being too small.

5.4.3.4 Input allocative efficiency

The next step of the analysis was the measurement of input allocative efficiency. The summary results for the production approach are shown in Table 5.11

Table 5.11 - *Input allocative efficiency for the production approach*

<i>Input allocative efficiency</i>	
Mean	84.9%
Standard deviation	12.1%
No. of efficient branches	12

Table 5.11 shows that the allocative efficiency is, on average, 85%. This value indicates the proportional reduction to technical efficient costs (i.e., the total cost after eliminating pure technical and scale inefficiencies) that would be achieved by adopting the right input mix in light of current prices.

The measure of allocative efficiency for an assessment based on the production approach indicates whether the operational costs and staff mix is appropriate for the current output levels of the branch, given the current relative salaries. The results of the efficiency analysis showed that the reallocation of staff between branches could bring substantial efficiency improvements to the network, as only twelve branches were identified as allocative efficient.

5.5 Managerial implications of the efficiency assessment

5.5.1 Comparative analysis of production and intermediation efficiency

Table 5.12 reports a summary of the efficiency measures from the production and intermediation approaches. The results of the intermediation approach are based on the

output set defined in terms of revenues. As the revenues are a function of the amount of money intermediated and interest received, this output set is the most comprehensive, and best suited for the comparison between operational and financial performance.

Table 5.12 – *Summary of the efficiency measures for the production and intermediation approaches*

	<i>Intermediation approach</i> <i>(output set based on revenues)</i>	<i>Production approach</i>
Cost efficiency	87%	69%
Input allocative efficiency	91%	85%
Technical efficiency	96%	81%
Pure technical efficiency	---	92%
Scale efficiency	---	88%

We start the discussion of the empirical results with the comparison of the **cost efficiency** measure for the production and intermediation approaches. Although the difference between the cost efficiency measures of both approaches appears to be quite substantial, some further analysis suggests that these results are consistent.

The main reason for the difference in the efficiency estimates of the two approaches is that the financial aspects of branches' business are only reflected in the intermediation approach (recall that interest costs and revenues are only included in this approach). In this network, interest costs account, on average, for 80% of total branch costs. Since all branches have virtually identical values of spread from deposits and loans (which result from back office activities made centrally at the bank), the value of the potential cost reductions (measured in percentage) associated with the elimination of intermediation inefficiency could not be large. Recall that the DEA efficiency estimate is a relative measure, and since relevant aspects of branches' intermediation activity are defined centrally, the intermediation efficiency spread detected by DEA should be small.

Conversely, the efficiency of the operational activity is mostly determined by local management and the quality of branch staff. Therefore, the efficiency spread of the production approach is larger, as could be expected.

The results above show that inefficiencies account, on average, for 31% of branch operational costs (e.g., (100-69)%, obtained from the production approach) and 13% of branches' total costs (e.g., (100-87)%, obtained from the intermediation approach).

In relation to **technical efficiency**, it is interesting to compare the efficiency spread observed in the branches from the Portuguese bank with the results of previous studies based on the DEA technique for banks in other countries¹³. Most of the studies reported in the literature were based on the production approach. The average technical efficiency measure of 81% observed in this network is similar to the result of the study by Sherman and Ladino (1995), that revealed an average efficiency value of 80% for an U.S. bank. A few other studies (e.g., Parkan, 1987; Vassiloglou and Giokas, 1990; Drake and Howcroft, 1994; and Athanassopoulos, 1997) found higher values of relative efficiency, around 90%. This indicates that the Portuguese branches may have greater scope for efficiency improvements through learning from the best practices observed within the same network than most branches analysed in other countries.

In relation to **scale efficiency**, the empirical findings of this case study are also consistent with earlier studies based on the production approach (e.g., Drake and Howcroft, 1994; Giokas, 1991). This suggests that banks incur a certain amount of additional operational costs by 'overbranching' in order to gain additional revenues from providing extra customer convenience. For instance, the presence of an additional branch may contribute

¹³ These international comparisons should be done with caution, as the nature of DEA is such that it does not allow direct comparisons to be made with different samples and different input-output definitions.

to attract new customers and provide improved convenience for the existing bank customers. As a result, the revenue generated may be greater than the cost associated with operating another branch, increasing the bank profitability.

Note that in the case of this network the revenue enhancements achieved can be substantial, as scale size does not affect branches' financial performance (as modelled by the intermediation approach). Profit maximising banks may therefore have branches that are below most productive scale size (MPSS) in operational terms, in order to gain additional revenues from the intermediation activity. The study by Berger *et al.* (1997) found a similar picture of branches' performance based on the use a parametric stochastic frontier method.

From the results discussed above it can be concluded that reducing the number of branches may not be advisable in terms of network profits. We cannot tell from our data how revenue would respond to reductions in the number of branches, but it is plausible that customers could leave the bank, making revenue losses greater than the operational cost savings from closures or consolidations between branches. Thus, the bank may well be scale efficient from an overall perspective, as given by the intermediation approach, even if most branches are below the MPSS from an operational perspective.

5.5.2 Implications of bank mergers in terms of branches' efficiency level

The results of this efficiency assessment have important implications for the planning of bank mergers. Cost savings from mergers can be achieved through improvements in branch efficiency in either of two ways. Firstly, closing inefficient branches can bring performance improvements by moving their business to efficient branches that are better managed. This requires the existence of branches that are substantially more efficient nearby, and with the right conditions to absorb the business of the closed branches. Also,

scale efficiency improvements can be achieved if the branch consolidations can bring the branches below MPSS to efficient scale without adding more scale diseconomies to the branches currently larger than MPSS.

The other way in which branch efficiency can be improved through mergers is if the other bank can bring superior management capacity at a central level, which may improve the network performance.

In relation to the first method of efficiency gains through mergers, the scale inefficiency detected in branches' operational activity suggests that there could be large savings from branch consolidations, but it is possible that they are not all achievable. This is due to the potential revenue losses associated with the reduction of customer convenience when closing branches. Indeed, closing large numbers of branches might bring about substantial losses of customers and revenues, which could make many of these closures unprofitable. The only circumstances in which there is not likely to be a significant loss of business from branch closures is where there are co-located small branches of the merged institutions, both well below MPSS. This depends on the geographic overlap of the networks, although branches from different banks are often close together.

In terms of the network analysed, in order to reduce significantly the scale inefficiencies detected, the branches with IRS would need to increase their scale size by more than 30%, on average. This could imply a large number of branch consolidations and reallocation of customer accounts between branches. This may bring some difficulties in reorganising the network, limiting the efficiency gains through this method.

In relation to the alternative channel of efficiency gains through mergers, consisting of better management at the bank level, the large dispersion of efficiency levels detected in this case study suggests that central management is not able to control fully the branches'

costs. Rather, the quality of local management appears to be quite important in determining branches' performance. Therefore, the branches' efficiency gains obtained through improved managerial capacity at a central level are likely to be quite modest, as local management is important in determining branch efficiency, limiting the role of bank level management.

In conclusion, the most significant cost reductions achievable through mergers might be through the consolidation of the back-office operations of the merged banks and branch closures where there are co-located small branches. Other types of efficiency improvements at the branch networks as a result of mergers are likely to be modest.

Previous bank-level studies often found that the efficiency gains from mergers were modest, and sometimes even non-existent, e.g., Berger and Humphrey (1992), Grifell-Tatjé and Lovell (1996), Rhoades (1993 and 1998), Sherman and Rupert (1999). The efficiency improvements eventually achieved may sometimes be offset by inefficiencies created elsewhere in the consolidated bank.

5.6 Efficiency and profitability

The aim of this section is to integrate efficiency and profitability measures in an overall framework for assessing bank branches' performance. The joint use of the efficiency and profitability measures can highlight the potential performance improvements that management might be able to explore, leading to higher profits. This analysis is based on the 'efficiency-profitability matrix' proposed by Dyson *et al.* (1990) and Boussofiane *et al.* (1991).

We start by analysing the relation between intermediation efficiency (with the output set based on revenues) and profitability. Figure 5.5 shows the 'efficiency-profitability

matrix' for the intermediation approach. The efficiency measure represented is cost-effectiveness. It was chosen because it is the most comprehensive efficiency measure, resulting from the aggregation of all efficiency components (e.g., technical efficiency, allocative efficiency and output mix efficiency).

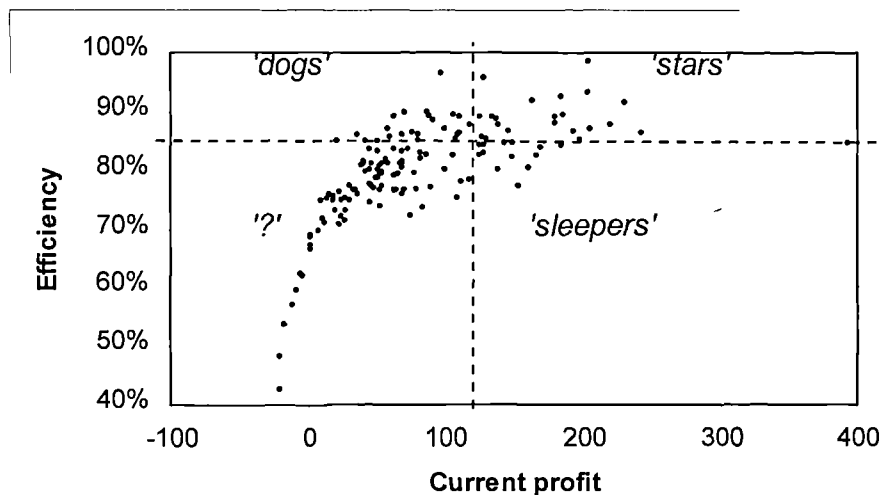


Figure 5.5 – *Intermediation efficiency versus profitability levels*

Figure 5.5 shows that high intermediation efficiency is associated with high profits. Most importantly, it becomes clear that the branches with the lowest efficiency values are also the least profitable.

The efficiency-profitability matrix was separated in four quadrants in order to identify different profiles of branches, as in Boussofiene *et al.* (1991). The precise boundary positions between the quadrants are subjective. In this analysis we have defined the cut-off point between high and low profits a value of 120 million escudos. This value ensured that about 25% of the branches were in the high profit area of the matrix, which accounted for about 50% of the network profits. In relation to the efficiency value, the cut-off point between high and low efficiency values corresponds to an efficiency level of 85%. This ensured that 30% of the branches were included in the high efficiency part of the matrix.

Branches located in the '*star*' quadrant ensure long-run economic viability, provided they maintain that position. These should be seen as the benchmarks for this network. The performance of these branches should be carefully monitored, as they are essential to the economic viability of the network. The scope for improvement in the profitability of these branches may be limited, as there is no empirical evidence that their activity can be improved significantly by emulating the 'best practices' observed in other branches of the same network. The benchmarking practices for these branches should be looked for in other bank branch networks. A closer matrix examination shows that 22 branches are 'stars'.

Branches in the '*dog*' quadrant are efficient but have low profits. These branches should be subject to a detailed assessment and eventually closed, as there is little scope for improvements, perhaps due to an unfavourable location. There are 22 branches located in this quadrant of the matrix.

Branches located in the '*question mark*' quadrant have the potential for both greater efficiency and profitability. These branches should improve their efficiency as an attempt to attain higher profits and move towards the 'star' quadrant. There are 86 branches located in the 'question mark' quadrant.

The '*sleepers*' are profitable, yet inefficient. Their profitability is likely to be a consequence of favourable environments, which can be exploited further by improvements in efficiency. These branches should be prime candidates for an efficiency improvement effort leading to greater profits. 14 branches are located in this part of the matrix.

The overall targets for the network in terms of efficient input and output levels (including the slack values) are shown in Table 5.13. The targets reported correspond to a sequential elimination of inefficiencies, starting from technical efficiency (TE), followed by input allocative efficiency (AE) and finally output mix efficiency (OME). Only the output targets corresponding to the elimination of output mix inefficiencies are reported, as the other efficiency assessments (i.e., TE and AE) were done with an input orientation, without involving output changes.

Table 5.13 - *Target input and output levels for the intermediation approach*

	<i>Initial values</i>	<i>TE targets</i>	<i>TE & AE targets</i>	<i>TE,AE & OME targets</i>
Non-interest costs	8865	8331	6525	4869
Interest costs from deposits	29693	28600	26965	16798
Interest costs from loans	13303	12742	13357	20919
Revenue from deposits	40492	---	---	29356
Revenue from loans	21460	---	---	31475
Value of Commissions	4199	---	---	5320
Total cost target	51861	49673	46847	42586

The greater cost reductions are attainable through the elimination of output mix inefficiencies. Comparing the cost-effectiveness targets with the initial output levels, it can be concluded that in order to maximise profits, the output mix should be reoriented towards customer loans and off balance sheet business instead of focusing on deposits, which is currently the main source of branches' revenue.

At the network level, the profitability increases associated with the elimination of the inefficiencies are shown in Figure 5.6. The values reported assume that total revenue cannot be changed, so that the target profits are achieved exclusively through cost reductions.

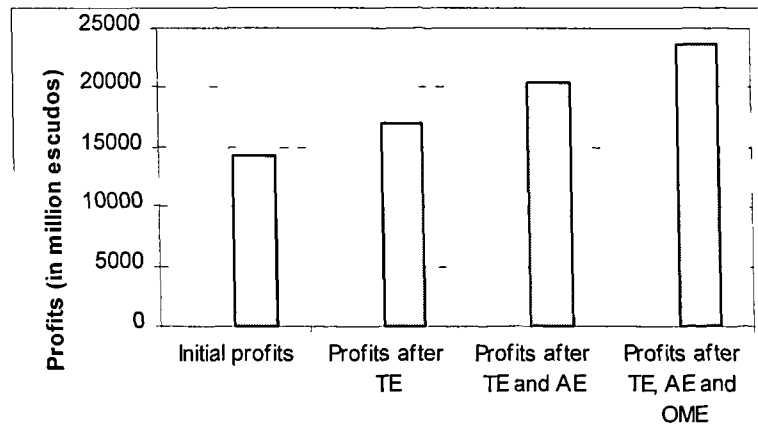


Figure 5.6 – *Profits levels achievable through the elimination of inefficiencies*

The next phase of the efficiency-profitability analysis focused on the relation between operational efficiency and profits. The corresponding 'efficiency-profitability matrix' is shown in Figure 5.7. The efficiency measure represented corresponds to cost efficiency.

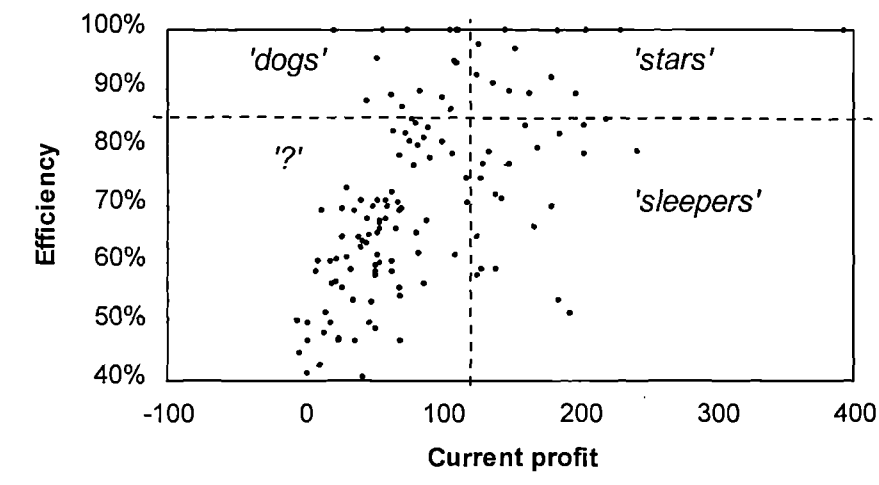


Figure 5.7 – *Operational efficiency versus profitability levels*

Figure 5.7 shows that although higher operational efficiency is associated with higher profits, there is a large variability in the efficiency levels of the most profitable branches. This indicates that high profits can be attained despite the existence of operational inefficiencies. The large efficiency spread pictured in Figure 5.7 suggests that there is scope for profitability improvements by eliminating operational inefficiencies.

The branches profile in terms of the operational activity can also be characterised based on their location in the matrix (e.g., ‘stars’, ‘dogs’, ‘sleepers’ or ‘question marks’), as done earlier for the intermediation approach.

The sources of operational inefficiency will be analysed in detail in the following chapters. To motivate the research of the following chapters, it is possible to say from the overall analysis made so far that the network profitability may be increased around 16% by attaining operational efficiency¹⁴. This would not require any changes to branches’ current output levels or to financial intermediation aspects such as spread of deposits and loans.

5.7 Summary and conclusions

From a conceptual point of view, this chapter addressed the following issues: (i) The use of a DEA model that can identify both input and output inefficiencies with a cost minimisation perspective; (ii) The integration of the production and intermediation approaches in the assessment of efficiency of financial institutions; (ii) The analyses of the impact of bank mergers in branches’ efficiency; and (iv) The integration of efficiency and profitability measurement in a general framework for performance assessment.

The cost-effectiveness model developed highlighted that the major source of branches’ inefficiency is related to the balance of the banking products sold. Adopting different output mixes could reduce significantly the branches’ total costs. This highlights the importance of using enhanced efficiency models that can simultaneously consider both the input and output sides of business.

¹⁴ This potential profitability increase was calculated considering that the cost at each branch can be reduced from the current level to the cost target obtained in the cost efficiency assessment (using model (2.20) in Chapter 2).

We have shown that the production and intermediation approaches are a powerful tool when applied jointly. This case study showed that most branches are smaller than efficient scale from an operational perspective, consistent with prior findings reported in other DEA studies. However, the profitability levels achieved for the entire network may justify some of this scale inefficiency detected in branches' operational activity, as scale size does not affect branches' financial performance (as suggested by the results of the intermediation approach). Additional branches provide extra convenience for the bank's customers, who may be willing to pay higher prices for the financial products in exchange for the extra convenience. Therefore, the elimination of other inefficiency sources (e.g., pure technical, input allocative and output mix inefficiencies) should be given pre-emptive priority.

The empirical results also suggest that cost savings in the branch networks as a result of mergers may be limited. To achieve savings from closing branches that are below efficient scale requires other branches that are also below efficient scale and geographically proximate to absorb the additional business coming from the closed branches. To achieve substantial savings from closing branches with other sources of inefficiencies (pure technical, input allocative or output mix inefficiencies) requires the presence of nearby branches that are efficient and can therefore do a better job at managing the business of the closed branches. Although the geographic overlap of networks from merged banks is likely to exist, as branches of different banks are often close together, if the branches to which the business is transferred are not as convenient for customers as the closed branches, the loss of market share and revenues may more than offset the cost savings.

The other possibility for branch efficiency improvements as a result of mergers could occur if a superior bank management takes over and improves the management of a poorly run network. The results discussed in this chapter suggest that local branch management is important in determining branch efficiency, which limits the role of bank-level management.

All these limitations are bolstered by the general finding in the literature that bank mergers do not have much effect on efficiency improvements leading to a substantial rationalisation of costs. The most significant cost reductions achievable through mergers might be through the consolidation of the back-office operations of the merged banks and branch closures where there are co-located small branches.

Another implication of the analysis reported in this chapter is that there are a number of ways in which banks can use branch efficiency measures, in conjunction with their own performance measurement system, to make their branch networks more efficient and profitable. The representation of branches within an 'efficiency-profitability matrix' can provide a useful characterisation of branches' performance profile. Branches located in different quadrants of the matrix require different performance improvement strategies in order to increase the bank's profitability. The efficiency assessment can also be used as a tool to quantify the potential profitability improvements and identify the input and output targets that would enable the achievement of the maximal profits.

Observation of the most efficient and least efficient branches can also help discover efficient and inefficient practices, respectively, that may be used to improve managerial policies and procedures. In addition, relative efficiencies may be used as an incentive or monitoring device for staff performance appraisal.

CHAPTER 6

Improving operational activity: Pure technical and scale efficiency issues

6.1 Introduction

The previous chapter developed an overall framework for the assessment of performance of financial institutions, integrating efficiency and profitability measures. Branches' efficiency was assessed under two perspectives, corresponding to the operational and intermediation activities. Most of the inefficiencies detected were associated with operational aspects, involving the proportionate overuse of resources (pure technical inefficiency), having a scale size that does not enable the attainment of maximal productivity (scale inefficiency) and having an improper mix of inputs in light of prices (input allocative inefficiency).

The purpose of this chapter is to explore in greater detail these operational inefficiencies. Only technical aspects of efficiency, defined in terms of physical production possibilities are addressed. The economic aspects associated with the choice of an optimal resource balance for cost minimisation are addressed in Chapter 7. In particular, this chapter focuses on the effect of scale size on branches' efficiency and on the identification of targets leading to improved performance.

The recent strategy of the bank used as a case study has focused on growth of business levels through mergers with other financial institutions. This required rationalisation of resources in existing branches and redeployment of surplus staff to new ones. The

general policy towards growth has been to open small branches with four members of staff. Thus, the relation between branches' scale size and operational performance is explored in greater detail in this chapter, in order to identify the optimal scale size for the existing and new branches.

This chapter also focuses on the development of target setting methods to improve efficiency. Practical issues about the DEA results and usage are discussed. Two alternative target-setting strategies were considered. One is concerned with the elimination of pure technical inefficiencies focusing on the identification of appropriate benchmarks for guiding the improvement efforts. The other concerns the attainment of branches' most productive scale size through the elimination of scale inefficiencies, with the minimal possible changes to branches' scale size.

This chapter is structured as follows. Section 6.2 describes the inputs and outputs used in the analysis described in this chapter. Section 6.3 explores the impact of branches' scale size on efficiency and productivity. Section 6.4 develops enhanced target setting methods, focusing on the identification of appropriate benchmarks for inefficient DMUs and on the elimination of scale inefficiencies avoiding significant changes to the DMUs scale size. Section 6.5 summarises and concludes.

6.2 Input and output measures

As the primary aim of this chapter was to explore further the effect of scale size on branches' operational performance, the analysis adopted the main structure of the production approach. The input and output sets were redefined with the objective of representing more accurately the impact of scale size on branches' business, as follows:

- Inputs :**
- Number of employees at the branch.
 - Operational costs (excluding staff costs).^{†1}
 - Floor space of the branch (in m²).
 - Number of external ATMs.
- Outputs:**
- Total value of deposits.[†]
 - Total value of loans.[†]
 - Total value of off balance sheet business.[†]
 - Number of general service transactions (done by staff).[†]
 - Number of transactions in external ATMs.
 - Number of all types of accounts at the branch.

In relation to the **input** set, the staff with different functions (e.g., branch/account managers, administrative/commercial staff and tellers) was aggregated and included in the model as a single variable, representing the (average) number of employees working at the branch during 1996. This aggregation enabled including other variables without losing the discrimination power of the DEA model. The operational costs variable refers to costs of supplies and services from other companies, commissions paid, and other costs. The floor space corresponds to the area available in each branch. The space associated with activities not directly related with branches' business was not included (e.g., training spaces, central archives, etc.). The number of external ATMs reflects the technology available in each branch for performing automated transactions.

Neither the technology within the branch (e.g., number of on-line terminals) nor the state of the premises has been included in the model because the equipment and branch image are fairly homogenous in all branches of the network analysed.

¹ The input and outputs marked with the symbol [†] are identical to those defined for the production approach in Chapter 5.

The **output** set differs from the one used in the previous chapter due to the inclusion of two additional variables, representing the number of accounts at the branch and the number of transactions made in ATMs. The purpose of the output variables defined is as follows. The number of transactions performed by branch staff includes an aggregation of 20 different types of general service transactions, (e.g., cheques processed, purchase of foreign currency, etc.), representing the workload of tellers and administrative/commercial staff. In order to account for more specialised and complex transactions, generally involving branch/account managers, (e.g., number of loans negotiated, pension funds set up, personal credits, etc.), the model included stock measures of their outcomes (i.e., value of deposits, value of loans and value of off balance sheet business). The number of accounts of all types in the branch reflects the ability to attract and maintain a large customer base at the branch. Finally, the number of transactions in ATMs represents the volume of services provided through automated means.

Table 6.1 displays the summary statistics of the inputs and outputs defined, including the mean, standard deviation (SD) and extreme values.

Table 6.1 - *Summary statistics of the input and output data*

<i>Inputs/Outputs</i>	<i>Mean</i>	<i>SD</i>	<i>Minimum</i>	<i>Maximum</i>
No. of employees	9.8	3.7	4	21
Operational costs	18.1	10.1	7.4	105.4
Floor space	294.4	168.3	80	966
No. of external ATMs	1.1	0.4	0	3
Total value of deposits	3662.1	3042.5	202.1	19080.2
Total value of loans	1214.8	1020.5	131.1	7104.8
Total value of off balance sheet business	1347.8	1268.7	23.4	11388.0
No. of general service transactions	92421.6	42509.2	16221	212566
No. of transactions in external ATMs	6193.7	3716.6	0	20136
No. of accounts	6645.9	4079.9	1115	21619

6.3 Effect of scale size on branches' operational activity

The impact of scale size on banking business is well documented at the bank level (see McAllister and Mcmanus, 1993). However, to date only a limited number of studies focused on branches' optimal scale size in terms of efficiency and productivity.

Giokas (1991) was the first study to use a VRS model for the analysis of bank branches' performance. The average scale efficiency found was high, and the majority of scale inefficient branches had increasing returns to scale (IRS). However, the analysis of the returns to scale nature reported in this paper had some flaws, since the existence of multiple optimal solutions to the DEA model was not accounted for.

The results of the study by Drake and Howcroft (1994) were very similar, i.e., the scale efficiency detected was high and most scale inefficient branches had IRS. This study reported that the optimal branch size in terms of the efficiency of the operational activity was 9 employees.

Schaffnit *et al.* (1997) also found high scale efficiency values, but a different picture in relation to branches' returns to scale. The majority operated under CRS and of the remaining branches most had DRS.

Finally, the study by Athanassopoulos (1998) found different returns to scale characteristics according to branches' activity profile.

The remainder of this section contributes to the clarification of the impact of branches' scale size on efficiency and productivity, based on the case study of the Portuguese bank.

6.3.1 Definition of branch scale size

As the major concern of the bank is the efficient use of resources, the measure of branches' scale size should be the variable that best reflects the resources level of the branches. As the personnel costs account for approximately 75% of the total operational costs, the variable 'number of employees' included in the DEA model was considered the most appropriate measure of branch scale size. Several concepts of scale size can be defined:

- Current scale size: Current number of employees working at the branch;
- Pure technical efficient scale size: Number of employees after the elimination of pure technical inefficiencies;
- Most productive scale size: Number of employees after the elimination of both pure technical and scale inefficiencies.

According to these scale size definitions, a branch might have to change its current scale size to attain pure technical efficiency. However, it is only considered scale inefficient if its pure technical efficient scale size does not coincide with its most productive scale size. Note that the scale efficiency notion, as well as the characterisation of a DMU as exhibiting increasing, decreasing or constant returns to scale can only be referred to DMUs free of pure technical inefficiencies, i.e., DMUs located on the VRS frontier. Figure 6.1 shows the distribution of branches by current scale size.

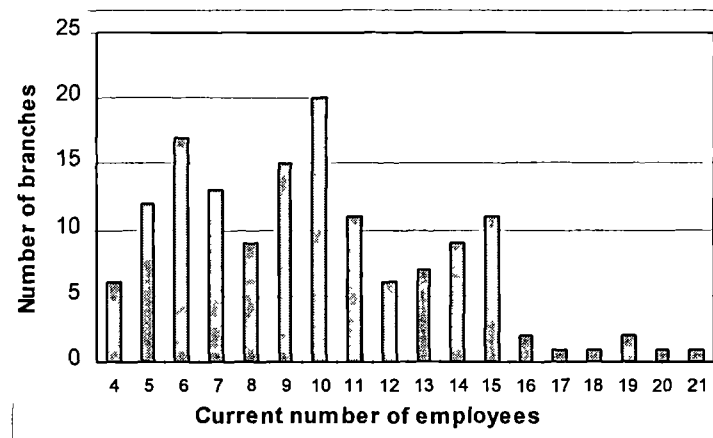


Figure 6.1 - *Distribution of branches by current scale size*

The average scale size of the branches is 10 employees. There are a few branches with a very large number of employees and six branches operating with only 4 employees.

6.3.2 Input versus output oriented assessments

The first step of the efficiency assessment consisted of the identification of the returns to scale properties of branches' production technology, based on the input-output set used in this chapter. This was made following the procedure proposed in Banker (1996).

The Kolmogorov-Smirnov test indicated that the branches operate under VRS². This conclusion is similar to the one obtained in the previous chapter, which indicates that the returns to scale characteristics of the production technology are not affected by the choice of variables used to represent branches' operational activity.

In order to explore in greater detail the relation between branch scale size and pure technical and scale efficiencies, both input and output orientations of the efficiency assessment were examined. Under VRS, input and output oriented efficiency assessments are fundamentally different concepts that can lead to different efficiency measures for inefficient branches.

Table 6.2 shows the average pure technical efficiency and scale efficiency for all branches in the sample, both for input and output oriented assessments.

Table 6.2 - *Pure technical efficiency (PTE) and scale efficiency (SE) results*

	<i>Mean PTE</i>	<i>Mean SE</i>
Input orientation	91.7%	87.8%
Output orientation	86.7%	93.1%

² The null hypothesis of identical inefficiency distributions under CRS and VRS evaluations was rejected with a p-value of 0.000.

Figure 6.2 and Figure 6.3 show the average pure technical and scale efficiencies of branches grouped by scale size, for input and output oriented assessments, respectively.

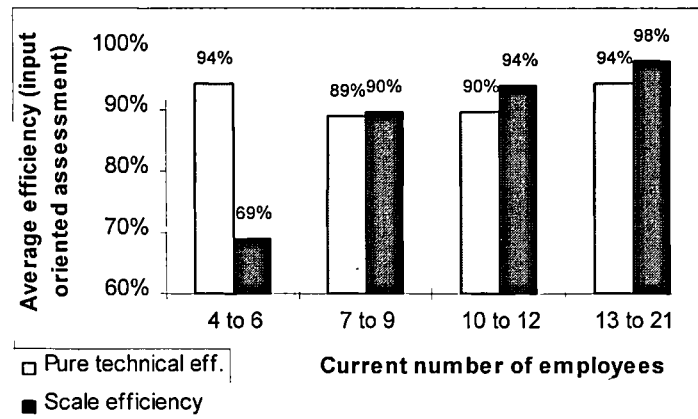


Figure 6.2 - *Efficiency components and branches' scale size (input oriented assessment)*

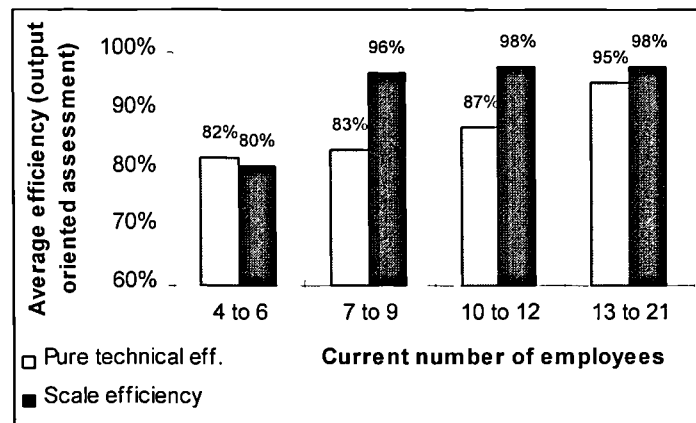


Figure 6.3 - *Efficiency components and branches' scale size (output oriented assessment)*

If an input orientation is adopted, consistent with the goals expressed by branch managers interviewed during the analysis of this case study, then the pure technical efficiency improvements should be primarily sought on branches with a number of employees between 7 and 12. These branches have the lowest value of pure technical efficiency, approximately 90%, as shown in Figure 6.2. However, after achieving pure technical efficiency the smaller branches would still exhibit significant scale inefficiencies. This indicates that the resulting branches' scale size does not enable the maximisation of productivity, due to the inherent returns to scale properties of branches' activity.

On the other hand, the adoption of an input orientation for the elimination of inefficiencies will generate a staff surplus from the rationalisation of resources. These employees can be relocated to open new branches. Although an input orientation is not the best for improving operational efficiency, it enables improving customer service through the convenience created by a larger network. This may bring additional revenues to the bank, which compensate the cost penalties of the scale inefficiencies in the operational activity, as discussed in the previous chapter.

If an output orientation is adopted, which assumes that there is still scope for the growth of branches' business, then the pure technical efficiency improvements should be primarily sought among smaller branches, with a number of employees between 4 and 9 (see Figure 6.3). Although the resulting activity levels free of pure technical inefficiencies would still have scale inefficiencies, their magnitude would be smaller than if an input orientation was adopted.

It should be noted that independently of the orientation chosen for the efficiency assessment, scale inefficiencies are almost non-existent in larger branches.

6.3.3 The components of the operational efficiency measure

Given the returns to scale nature of branches' operational activity, the technical efficiency measure corresponds to the aggregation of pure technical and scale efficiency components.

The relative magnitude of these two components was further investigated in this section. A direct comparison of their values can be misleading as they represent proportional changes corresponding to different operating positions within the production possibility set.

This issue can be pictorially illustrated in Figure 6.4, which represents the input and output levels of DMUs A to F, using a single input to produce a single output.

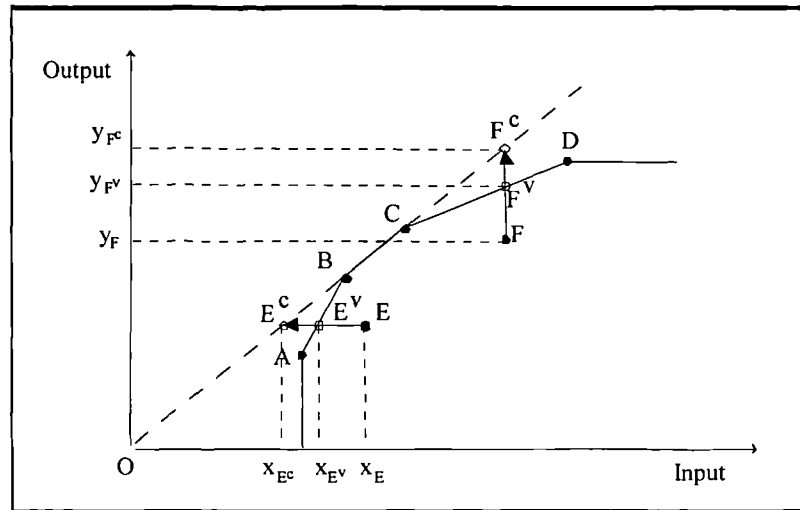


Figure 6.4 - Efficiency components under VRS

The line spanned by OB is the CRS efficient frontier, whilst the segments that link DMUs A, B, C and D form the VRS efficient frontier.

With an **input orientation**, the measures of TE, PTE and SE for DMU E can be obtained

as follows: $TE = \frac{x_{E^c}}{x_E}$, $PTE = \frac{x_{E^v}}{x_E}$, $SE = \frac{x_{E^c}}{x_{E^v}}$. Note that TE is equal to the product of

PTE and SE.

From the formulas above, it becomes clear that the magnitudes of PTE and SE should not be directly compared, as PTE corresponds to a proportion of the DMU's *current input level* (x_E), whilst SE corresponds to a proportion of the *input level free of pure technical inefficiency* (x_{E^v}).

In order to allow the direct comparison of the magnitudes of pure technical inefficiency and scale inefficiency, an additive decomposition of *technical inefficiency* is proposed,

where the two components are referred to the same basis. For an input oriented assessment these measures are defined as follows:

- Technical inefficiency (TI): $TI = 1 - TE = \frac{x_E - x_{E^c}}{x_E}$.
- Pure technical inefficiency (PTI): $PTI = 1 - PTE = \frac{x_E - x_{E^v}}{x_E}$.
- Adjusted scale inefficiency (SI): $SI = PTE - TE = \frac{x_{E^v} - x_{E^c}}{x_E}$.

All the above measures correspond to the proportion of the *current input level* that is wasted due to inefficiencies. With these definitions of inefficiency, technical inefficiency is equal to the sum of pure technical inefficiency and *adjusted* scale inefficiency.

In terms of the comparison of magnitude between these inefficiency components, if the pure technical inefficiency measure is greater than the *adjusted* scale inefficiency measure, the major source of the DMU's inefficiency is pure technical. Conversely, if the scale inefficiency measure is greater than the pure technical inefficiency measure, the major source of inefficiency is due to operating at a wrong scale size.

A similar decomposition can be made for an **output-oriented** assessment. For illustrative purposes, consider DMU F in Figure 6.4. The efficiency measures are as

follows: $TE = \frac{y_F}{y_{F^c}}$, $PTE = \frac{y_F}{y_{F^v}}$, $SE = \frac{y_{F^v}}{y_{F^c}}$. Note that TE is equal to the *product* of PTE and SE.

Again, the magnitude of PTE and SE should not be directly compared, as PTE refers to the proportion of *maximal output under VRS* (y_{F^v}) that is currently achieved, whereas SE refers to the proportion of *maximal output under CRS* (y_{F^c}) that is achievable when operating on the VRS frontier.

To enable a direct comparison between pure technical and scale inefficiency components, the following formulas should be used:

- Technical inefficiency (TI): $TI = 1 - TE = \frac{y_{Fc} - y_F}{y_{Fc}}$.
- Adjusted pure technical inefficiency (PTI): $PTI = SE - TE = \frac{y_{Fv} - y_F}{y_{Fc}}$.
- Scale inefficiency (SI): $SI = 1 - SE = \frac{y_{Fc} - y_{Fv}}{y_{Fc}}$.

These measures represent the proportion of the maximal output that is wasted due to inefficiency. For an output-oriented assessment, technical inefficiency is equal to the sum of *adjusted* pure technical inefficiency and scale inefficiency.

Table 6.3 reports the technical inefficiency measure and its components for the branch network under analysis, both for input and output orientated assessments.

Table 6.3 – *Components of technical inefficiency*

	<i>Mean TI</i>	<i>Mean PTI</i>	<i>Mean SI</i>
Input orientation	19.3%	8.3%	11.0%
Output orientation	19.3%	12.4%	6.9%

With an input orientation, the scale inefficiency component is greater than the pure technical inefficiency component. Therefore, after the rationalisation of resources usage through the elimination of pure technical inefficiency, most branches will end up operating at a scale size that does not enable maximal productivity.

However, if an output orientation is chosen, most inefficiency is of pure technical nature, and can be eliminated with managerial improvement efforts. The resulting scale inefficiency (corresponding to operating at a scale size that does not enable maximal productivity) would be much smaller than if the efficiency improvement was made with an input orientation.

6.3.4 Returns to scale characterisation

The returns to scale properties of the DMUs are determined by their position on the VRS frontier. For DMUs not operating on the frontier, their returns to scale can only be determined after the elimination of pure technical inefficiency through the projection towards the VRS frontier.

However, depending on the direction chosen for the projection, the DMUs can end up on different facets of the frontier, where the classification regarding returns to scale can be contradictory. A characterisation of the DMUs' returns to scale that overcomes this issue is explored next.

Consider Figure 6.5, which shows a set of DMUs using one input to produce a single output. The DMUs A, B, C and D form the VRS efficient frontier. The segments AB, BC and CD represent the increasing returns to scale (IRS), constant returns to scale and decreasing returns to scale (DRS) subsets of the efficient frontier, respectively.

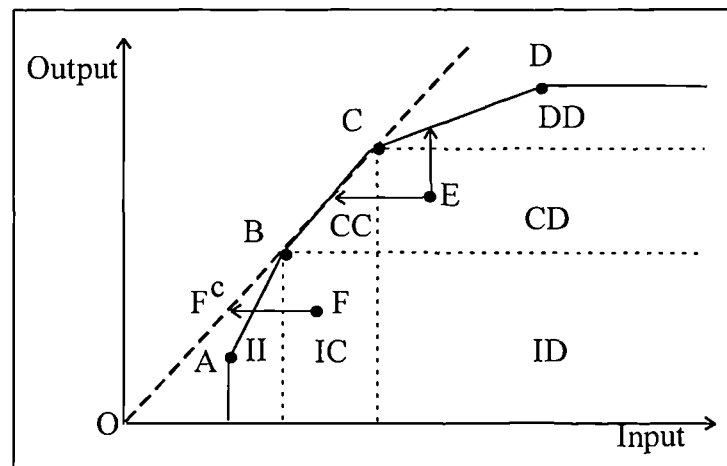


Figure 6.5 – Returns to scale characterisation

The returns to scale properties of DMU E will be analysed in greater detail. Adopting an input orientation for the elimination of pure technical inefficiencies, this DMU would be projected onto the CRS part of the frontier. Consequently, it would be classified as a

CRS DMU. Conversely, with an output orientation, the projection towards the frontier would result in a position under DRS. This example shows that the returns to scale characterisation of an inefficient DMU depends both on its location on the production possibility set and the direction of the projection towards the frontier.

Assuming that the DMUs cannot be projected to positions on the VRS frontier associated with an increase of input usage or a decrease in the outputs produced (e.g., for DMU E, the only projections allowed correspond to the shaded area in Figure 6.5), the following returns to scale characterisation applies. It corresponds to a partition of the PPS in 6 regions, with different characteristics regarding returns to scale.

- **II group:** The DMUs exhibit IRS independently of the orientation chosen for the assessment.
- **CC group:** The DMUs exhibit CRS independently of the orientation chosen for the assessment.
- **DD group:** The DMUs exhibit DRS independently of the orientation chosen for the assessment.
- **IC group:** The DMUs can exhibit IRS or CRS, depending on the orientation chosen for the assessment.
- **CD group:** The DMUs can exhibit CRS or DRS, depending on the orientation chosen for the assessment.
- **ID group:** The DMUs can exhibit IRS, CRS or DRS, depending on the orientation chosen for the assessment.

The partition of the PPS in six regions was originally proposed by Fare, Grosskopf and Lovell (1994), without empirical applications. Fukuyama (1996) used this classification to explore the differences in returns to scale properties of credit associations in Japan, resulting from input-oriented versus output-oriented assessments.

For the branch network analysed in this thesis, the number of branches in each of the returns to scale groups defined is as follows: 69 belong to group **II**, 34 to **CC**, 18 to **DD**,

9 to **IC** and 14 to **ID**. None of the existing branches belongs to group **CD**. Figure 6.6 illustrates these results. Figure 6.7 shows the returns to scale characteristics of branches grouped by scale size.

Distribution of branches in the returns-to-scale groups

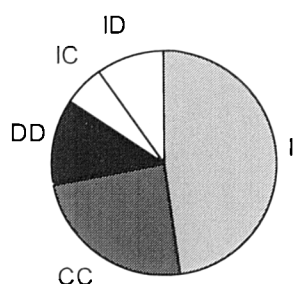


Figure 6.6 – Characterisation of branches' returns to scale

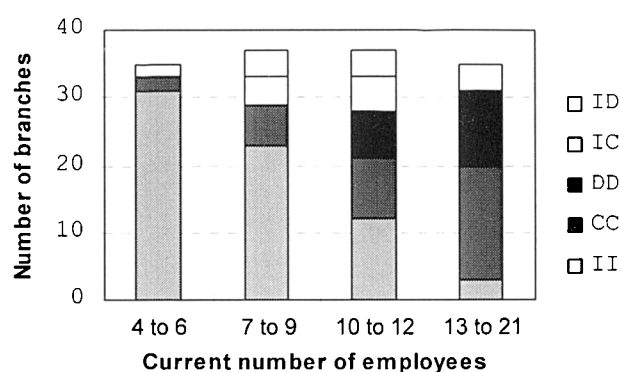


Figure 6.7 - Returns to scale and branches' scale size

Whilst branches located in the II region have up to fifteen employees, all branches located in the DD region have ten or more employees. The branches with constant returns to scale (i.e., CC group), and thus with optimal scale size, are predominantly in the largest size groups.

The analysis of the returns to scale characteristics of individual branches is important to determine the best orientation for efficiency improvement for each branch. The branches in the II and IC groups should focus on increasing business levels in order to improve the productivity of current resources. The branches located in the DD and CD regions should

focus on the rationalisation of resources, improving efficiency with an input orientation. Branches in the ID region should adopt the orientation for which the scale efficiency level after the attainment of PTE is higher. Based on this criterion, from the fourteen branches in the ID group, four should adopt an input orientation and the other ten should follow an output orientation. Branches in the CC region can adopt either an input or an output orientation for the efficiency improvements.

6.4 Target setting

An important aspect of a DEA efficiency assessment is to determine the input-output targets that would render the DMUs efficient. Two different target-setting scenarios were considered in this chapter:

- The first scenario concerns the attainment of pure technical efficiency. The DEA assessment focused on the identification of appropriate benchmarks for guiding the improvement efforts of inefficient DMUs.
- The second scenario concerns the attainment of pure technical and scale efficiency. The assessment focused on the identification of the scale size with maximal productivity closest to a DMU's scale size after the elimination of pure technical inefficiency.

In order to address some issues relating to efficiency improvements under both scenarios, enhanced DEA models and methods were developed. These are described in the next section.

6.4.1 Targets with selection of appropriate benchmarks

6.4.1.1 Motivation

When assessing a DMU j_0 , the standard VRS model (Banker *et al.*, 1984) can use any of the efficient DMUs to build the 'composite' peer DMU against which DMU j_0 is

compared³, provided the envelopment constraints are satisfied. These constraints impose that the convex combination of inputs (outputs) corresponding to the 'composite' peer must be at or below (above) the levels observed in the DMU j_0 under assessment, and no extrapolation is allowed, (i.e., the sum of the weights used to construct the 'composite' peer based on the input-output levels of efficient DMUs must be equal to one).

It is possible that some of the peers used to build the 'composite' DMU have a scale size very different from that of the assessed DMU j_0 . In these cases, the peers obtained as by-products of the DEA assessment may not be suitable benchmarks for DMU j_0 , as their size, and consequently the operating practices, may not be easily transferable to DMU j_0 .

In relation to the case study analysed in this thesis, as a branch's scale size can significantly affect efficiency, the standard VRS model (Banker *et al.*, 1984) was modified in order to preclude from the peer set branches that are either too large or too small to be considered benchmarks for the assessed branch.

The general formulation of the DEA model enabling a restricted choice of peers is described in the next section.

6.4.1.2 DEA model with restricted peers

One of the advantages of this enhanced DEA model is that it avoids the need to define clusters of homogeneous DMUs prior to the DEA analysis. It only requires defining a criteria regarding the acceptable difference between the input-output levels of the DMU j_0 and its peers. By including additional restrictions to the envelopment

³ This interpretation of the DEA efficiency score is based on the envelopment formulation (see models 2.15 and 2.16 in Chapter 2).

formulation of the standard VRS models (i.e., (2.15) or (2.16) in Chapter 2), the ‘composite’ peer against which DMU j_0 is compared will be based on a convex combination of DMUs that satisfy the similarity criteria defined.

The general formulation of this DEA model with restricted peers is presented in (6.1). Only an input-oriented model is shown. The output-oriented formulation can be obtained by adding the restriction regarding the acceptable peers (i.e., restriction (6.1a)) to a standard output-oriented VRS model in envelopment form.

DEA input oriented model with restricted peers	(6.1)
$\text{Min } \hat{e}_{j_0}^P = \theta_0^P - \varepsilon \sum_{r=1}^s s_r - \varepsilon \sum_{i=1}^m s_i$ $\text{s.t. } x_{ij_0} \theta_0^P - \sum_{j=1}^n x_{ij} \lambda_j - s_i = 0, \quad i = 1, \dots, m$ $\sum_{j=1}^n y_{rj} \lambda_j - s_r = y_{rj_0}, \quad r = 1, \dots, s$ $\sum_{j=1}^n \lambda_j = 1$ $\lambda_j = 0 \text{ if } \begin{cases} x_{ij} \notin [x_{ij_0} \pm \delta_i] \forall i \\ y_{rj} \notin [y_{rj_0} \pm \rho_r] \forall r \end{cases} \quad j = 1, \dots, n \quad (6.1a)$ $\lambda_j, s_i, s_r \geq 0, \quad \forall j, i, r$	

The notation used is as follows. $\hat{e}_{j_0}^P$ is the efficiency estimate for the input oriented VRS model with restricted peers. δ_i is the tolerance regarding the similarity between the value of input i at DMU j_0 and at its peers. ρ_r is the tolerance regarding the similarity between the value of output r at DMU j_0 and at its peers. All other notation is as used in the previous chapters.

Since the peer restricted model (6.1) results from the addition of the restriction (6.1a) to the standard VRS model, the efficiency rating of the modified model will always be greater or equal to the pure technical efficiency rating of the standard VRS model. Furthermore, the tighter the tolerances imposed to the input and output levels of the peers the higher will be the efficiency scores of the modified model. In the limit, when all tolerances are zero, the only peer allowed for comparison will be the DMU j_0 . Under such circumstances all efficiency scores will be equal to one.

Therefore, this model cannot alone be used for an efficiency assessment. It should be considered a tool for facilitating the improvement of performance of inefficient DMUs. Using model (6.1) each inefficient DMU is initially motivated to reach a ‘best-practice’ in a group of branches of similar characteristics (these peers must satisfy the similarity criteria initially defined). This corresponds to a movement towards an “intermediate frontier”, corresponding to a local optimal, such that comparability between the DMU assessed and the benchmarks identified in the DEA assessment is guaranteed. The improvement targets can be gradually adjusted by relaxing the similarity criteria using larger tolerances, until the DMUs are directed to the overall best practice frontier, whose peers and targets are obtained from the standard VRS model.

For an input oriented assessment, the initial phases of this process consist of implementing the target input levels obtained from model (6.1). The final stage corresponds to improvements whose targets are equal to those obtained using a standard VRS model (i.e., model (2.15) in Chapter 2).

This process of implementing the targets gradually may be beneficial for many organisations. Major changes to operating procedures may give rise to upheavals that can

be traumatic. Thus, implementing the changes with a few small steps can be a better method to enhance performance through an effort of continuous improvement.

6.4.1.3 Empirical results

In the case of the bank branches under analysis, the only type of similarity imposed between branch j_0 and its peers concerned the scale size, i.e., the peers for branch j_0 could not differ in scale size by more than two employees (i.e., $\delta_{i_{\text{emp}}} = 2$, with i_{emp} being the input representing the number of employees of a branch).

As explained in the previous section, the use of model (6.1) led to an increase of the efficiency measure, and to less demanding input reduction targets. The average efficiency obtained from model (6.1) was 97% (with the restriction that peers could not differ from the branch assessed by more than two employees), whereas the value for the standard VRS model was 92%. The targets corresponding to the assessment with peers restricted are shown in Table 6.4.

Table 6.4 - Targets for the branch network with restricted peers

<i>Inputs/Outputs</i>	<i>Observed</i>	<i>Targets</i>	<i>% change</i>
No. of employees	1408	1369	97%
Operational costs	2600	2366	91%
Floor space	42394	37601	89%
No. of external ATMs	160	159	100%
Total value of deposits	527344	562135	107%
Total value of loans	174926	184765	106%
Total value of off balance sheet business	194078	211376	109%
No. of general service transactions	13308714	14088937	106%
No. of transactions in external ATMs	891898	980568	110%
No. of accounts	957011	1003508	105%

The aggregate results for the network show that the inputs used least efficiently are floor space and operational costs, which could be reduced to about 90% of current levels. However, the floor space may be difficult to change in existing branches, so this

information should mainly be used as a guideline for planning new branches. As a consequence of the restriction imposed to the scale size of the peer branches, the reduction suggested for the number of employees is rather small. However, at the network level the number of employees can still be reduced by approximately 3%, which corresponds to a surplus of 39 employees. In relation to the number of ATMs, these must be treated as integer units. The results indicate that only one branch in this network could reduce its number of external ATMs from 3 to 2. It is possible that this ATM could be more efficiently utilised if it were relocated to another branch. Overall, these input reductions could still support an increase between 5% and 10% to the branches' business level.

6.4.2 Most productive scale size targets

6.4.2.1 Motivation

The targets derived previously aim to eliminate pure technical inefficiency by rationalising the input usage. However, scale inefficiencies would still prevail. In order to explore further improvements in DMUs' performance, an alternative set of targets that eliminates both pure technical and scale inefficiencies is described in this section. These targets are based on the notion of the most productive scale size (MPSS) introduced by Banker (1984).

The estimation of the MPSS for a DMU seeks to obtain the scale size that maximises its productivity. In order to maximise productivity, a DMU should increase its scale size if IRS is prevailing, and decrease the scale size if DRS is prevailing.

In order to illustrate how the MPSS targets are obtained, consider again Figure 6.5. The MPSS based method would re-scale the CRS target for DMU F from its projection at F^C by expanding it to point B or C^4 . The advantage of this scaling is that the resulting target has a scale of operation more comparable to existing efficient DMUs with constant returns to scale.

However, in the presence of multiple optimal solutions to the DEA model, the MPSS for a DMU may not be unique, as illustrated for point F above, where both DMUs B and C are operating at a scale size with maximum productivity. Furthermore, maximum productivity can also be achieved through the projection of DMU F to any point on the frontier defined by the segment BC.

This section develops a method to choose a unique MPSS among the multiple MPSS targets that may exist for a given DMU. It is considered that the MPSS target should be as close as possible to the DMUs' pure technical efficient target. Thus, the DMUs with IRS are set the smallest feasible MPSS target (e.g., point B in Figure 6.5) and DMUs with DRS are set the largest feasible MPSS target (e.g., point C in Figure 6.5). The DMUs with CRS are already at the MPSS and so no further movements along the frontier are needed. The method used to obtain the MPSS targets according to the above criteria is detailed next.

⁴ Note that under VRS the target at F^C is not achievable, as it is beyond the boundary of the PPS defined by the segments between A, B, C and D. The expression to obtain a MPSS target based on the use of the DEA model is shown in Chapter 2, formula (2.18).

6.4.2.2 Method to derive the MPSS target closest to the DMUs' PTE scale size.

The description of this method considers that the elimination of pure technical inefficiency is done with an input orientation. The generalisation of the method to an output orientation is straightforward.

To start, suppose that an optimal solution ($\theta_0^{\text{original}}$) with a scaling factor⁵ smaller than one

($\Lambda^1 = \sum_{j=1}^n \lambda_j^* < 1$) has been obtained from a CRS input oriented envelopment model (i.e.,

model (2.9) in Chapter 2). Given the existence of an optimal solution to the CRS model with $\Lambda^1 < 1$, the returns to scale at DMU j_0 are either increasing (if $\Lambda^1 < 1$ in all multiple optimal solutions) or constant (if $\Lambda^1 = 1$ in any optimal solution), according to the criteria defined in Chapter 2, Table 2.2. Therefore, the identification of the returns to scale nature at DMU j_0 requires exploring the multiple optimal solutions that may exist to model (2.9). In addition, the existence of multiple optimal solutions can lead to several alternative MPSS targets for the DMU j_0 under assessment.

In order to identify the nature of the returns to scale at the DMU j_0 , (whose original optimal solution to model (2.9) had a scaling factor smaller than one), and choose an adequate MPSS target, the following model is solved to explore the alternative optimal solutions to model (2.9), see Banker, Chang and Cooper (1996):

⁵ The scaling factor was defined in Chapter 2, section 2.4.4.

<p>DEA model to explore alternative optimal solutions to a CRS input oriented envelopment model (DMU j_0 with either IRS or CRS).</p>	(6.2)
$\text{Max } \sum_{j=1}^n \hat{\lambda}_j + \epsilon \sum_{r=1}^s \hat{s}_r + \epsilon \sum_{i=1}^m \hat{s}_i$ $\text{s.t. } \sum_{j=1}^n x_{ij} \hat{\lambda}_j + \hat{s}_i = x_{ij_0} \theta_0^{\text{original}}, \quad i = 1, \dots, m$ $\sum_{j=1}^n y_{rj} \hat{\lambda}_j - \hat{s}_r = y_{rj_0}, \quad r = 1, \dots, s$ $\sum_{j=1}^n \hat{\lambda}_j \leq 1$ $\lambda_j, \hat{s}_i, \hat{s}_r \geq 0, \quad \forall j, i, r$	

In the model above, all variables are identified with the symbol $\hat{\cdot}$ and $\theta_0^{\text{original}}$ is equal to the original optimal solution to model (2.9) for DMU j_0 . Model (6.2) searches among all alternative optimal solutions to model (2.9) the one with the highest value of the scaling factor $\hat{\Lambda}^1$.

In relation to the returns to scale nature at DMU j_0 , if at the optimal solution to model (6.2) the scaling factor ($\hat{\Lambda}^{1*} = \sum_{j=1}^n \hat{\lambda}_j^*$) is smaller than one then DMU j_0 has IRS.

Alternatively, if the scaling factor is equal to one, then DMU j_0 has CRS.

To obtain the MPSS target closest to the PTE target of DMU j_0 , the scaling factor obtained at the optimal solution to (6.2) is used. Note that this scaling factor is as high as possible if the DMU j_0 has IRS or equal to one if it has CRS. Formula (6.3) gives a unique MPSS target for each DMU, satisfying the condition above. This ensures that the DMUs with IRS are set the smallest possible MPSS target. The symbol $\hat{\cdot}$ identifies the variables from model (6.2).

$$(x_{ij_0}^{MPSS}, y_{rj_0}^{MPSS}) = \left(\frac{x_{ij_0} \theta_0^{\text{original}} - \hat{s}_i^*}{\hat{\Lambda}^{I*}}, \frac{y_{rj_0} + \hat{s}_r^*}{\hat{\Lambda}^{I*}} \right) \quad (6.3)$$

So far we have analysed the case of DMUs with either IRS or CRS. To complete the analysis, consider the case of an optimal solution ($\theta_0^{\text{original}}$) to the CRS input-oriented envelopment model (2.9) with a scaling factor greater than one ($\Lambda^I = \sum_{j=1}^n \lambda_j^* > 1$). In this case, the returns to scale at DMU j_0 are either decreasing (if $\Lambda^I > 1$ in all multiple optimal solutions) or constant (if $\Lambda^I = 1$ in any optimal solution), according to the criteria defined in Chapter 2, Table 2.2.

In order to identify the nature of the returns to scale at the DMU j_0 , (whose optimal solution to model (2.9) had a scaling factor greater than one), and choose an adequate MPSS target, the following model is solved to explore the existence of alternative optimal solutions to model (2.9), see Banker, Chang and Cooper (1996):

DEA model to explore the alternative optimal solutions to a CRS input oriented envelopment model (DMU j_0 with either DRS or CRS). (6.4)	
$\text{Min } \sum_{j=1}^n \hat{\lambda}_j - \varepsilon \sum_{r=1}^s \hat{s}_r - \varepsilon \sum_{i=1}^m \hat{s}_i$	
$\text{s.t. } \sum_{j=1}^n x_{ij} \hat{\lambda}_j + \hat{s}_i = x_{ij_0} \theta_0^{\text{original}}, \quad i = 1, \dots, m$	
$\sum_{j=1}^n y_{rj} \hat{\lambda}_j - \hat{s}_r = y_{rj_0}, \quad r = 1, \dots, s$	
$\sum_{j=1}^n \hat{\lambda}_j \geq 1$	
$\hat{\lambda}_j, \hat{s}_i, \hat{s}_r \geq 0, \quad \forall j, i, r$	

The MPSS target closest to the pure technical efficient target of DMU j_0 can be obtained using the scaling factor (i.e., $\hat{\lambda}^{t*} = \sum_{j=1}^n \hat{\lambda}_j^*$) obtained at the optimal solution to model (6.4). Note that this scaling factor is as small as possible if DMU j_0 has DRS or one if it has CRS. A unique MPSS target satisfying the above criteria is also obtained using formula (6.3), considering this time that the symbol $\hat{\cdot}$ identifies the variables from model (6.4). This ensures that the DMUs with DRS are set the largest possible MPSS target.

6.4.2.3 Empirical results

Table 6.5 shows the aggregate targets for the branch network that eliminate both pure technical and scale inefficiencies through the adoption of the MPSS target closest to the branches' pure technical efficient target. These results were obtained using an input orientation for projecting the inefficient DMUs to the VRS frontier.

Table 6.5 – MPSS targets for the branch network

<i>Inputs/Outputs</i>	<i>Observed</i>	<i>MPSS targets</i>	<i>% change</i>
No. of employees	1408	1479	105%
Operational costs	2600	2491	96%
Floor space	42394	35504	84%
No. of external ATMs	160	159	100%
Total value of deposits	527344	716081	136%
Total value of loans	174926	228033	130%
Total value of off balance sheet business	194078	272782	141%
No. of general service transactions	13308714	17542722	132%
No. of transactions in external ATMs	891898	1249024	140%
No. of accounts	957011	1233512	129%

The MPSS targets, analysed at the network level, indicate that an increase of approximately 5% in the total number of employees, keeping the operational costs at their current level, could support an increase of more than 30% to the business levels. This corresponds to the efficient operation both in managerial (i.e. pure technical) and scale terms.

6.5 Summary and conclusions

This chapter analysed the impact of scale size on branches' operational efficiency and productivity. It was found that branches' operational activity exhibits variable returns to scale. A major contribution of this chapter was the use of a method that can provide an objective characterisation of the returns to scale nature of all branches, including those not operating on the frontier of the PPS. This enabled the identification of the best direction (i.e., input or output orientation) for the elimination of branches' pure technical inefficiencies.

In the case of the network analysed, the achievement of pure technical efficiency should follow an output orientation for most branches. From the 90 branches with pure technical inefficiencies, 73 should adopt an output orientation for the elimination of inefficiencies (i.e., 54 from the II group, 9 from the IC group and 10 from the ID group). The increase in business levels keeping the current resource usage would project these branches onto a position in the frontier of the production possibility set closer to the highest productivity levels than if an input orientation was adopted. Conversely, the elimination of inefficiency with an input orientation could be adopted for the remaining 27 branches (i.e., 13 from the DD group and 14 from the ID group). This would generate a surplus of resources that could support the opening of new branches, in line with the bank growth strategy.

This chapter also explored the consequences of adopting the same orientation for the elimination of pure technical inefficiency in all branches. A method to compare the relative magnitude of the two components of technical inefficiency, corresponding to pure technical inefficiency and scale inefficiency, was proposed. The results of the analysis of this branch network indicated that a recommended strategy for the network

growth would be the expansion of branches' business keeping the existing input levels. This would enable the elimination of operational inefficiencies whilst bringing the branches closer to the scale size with maximal productivity. However, if the local business potential does not support an increase in business at existing branches, then an input orientation may have to be adopted, which will result in lower productivity levels at many branches due to their resulting small scale size.

The identification of appropriate targets for inefficient DMUs is also a crucial issue of any efficiency assessment. This chapter enhanced the DEA method by considering practical issues relating to the implementation of the DEA results.

In order to strengthen the benchmarking capacity of DEA, a new model that emphasises the selection of appropriate peers for inefficient DMUs was developed. This model restricts the peers of the DMU under assessment according to a similarity criteria defined *a-priori* by the decision-makers. This ensures that the operating practices of the benchmarks identified by DEA are easily transferable to the DMU under assessment.

The target-setting process was extended further in this chapter to enable the achievement of the DMUs' most productive scale size. In order to keep branches' scale size as close as possible to the pure technical efficient scale size, a method to choose between the alternative MPSS targets for each branch was developed. It consisted of choosing the smallest MPSS for branches with IRS and the largest MPSS for branches with DRS. This ensures an elimination of inefficiencies as smooth as possible for the organisation, as it requires less effort in training and relocation of resources. Also, at the branch level, keeping the scale size as close as possible to the VRS target enables the attainment of maximal productivity involving the smallest possible changes to the operating practices.

CHAPTER 7

Cost efficiency under different price scenarios

7.1 Introduction

This chapter focuses on the assessment of cost efficiency in branches' operational activity. Cost efficiency evaluates the ability to produce current outputs at minimal cost. This requires input and output quantity data as well as input prices. The usual assumption underlying cost efficiency (CE) measurement is that input prices are fixed and known at each DMU, and possibly different among DMUs. In a recent discussion of cost efficiency Cooper *et al.* (1996) note that CE as usually defined can be of limited value in actual applications because of data requirements and unjustifiable assumptions. *Exact knowledge of prices is difficult and prices may be subject to variation in short periods.* Motivated by the context of bank branches' activity, the purpose of this chapter is to enhance CE measurement to account for alternative price scenarios that may exist in actual applications.

The first scenario analysed considers that prices are fixed at the DMU level. This corresponds to a short-term analysis, where adjustments to prices are not possible. In the case where the prices are known at each DMU, as assumed earlier in the thesis, the assessment may follow the approach described by Farrell (1957) and operationalised by Fare *et al.* (1985). However, in some situations the prices are unknown at the DMU level, and only the maximal and minimal price bounds can be estimated. In order to account for this price uncertainty, an enhanced CE measurement method is proposed.

The second scenario allows for price changes at the DMU level. This scenario is particularly relevant for assessments comprising a longer time period, such that the decision-maker (DM) can adjust both input levels and prices when implementing the performance improvement targets. In such cases, the targets obtained from the cost efficiency assessment represent the optimal combination of input and price levels that can minimise costs whilst adequately reflecting the existing trade-offs between the inputs at each DMU.

The alternative price scenarios considered in this chapter for the estimation of cost efficiency are summarised in Figure 7.1.

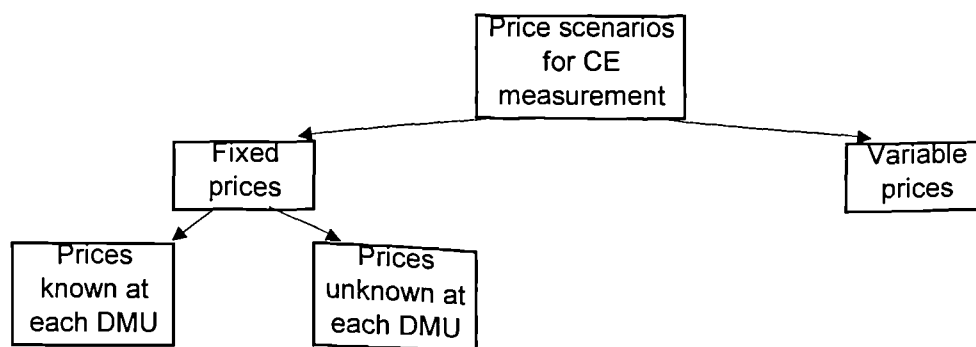


Figure 7.1 – Price scenarios considered for cost efficiency assessment

The structure of this chapter is as follows. The measurement of cost efficiency with fixed prices is discussed in section 7.2. Both scenarios of known prices and price uncertainty at the DMU level are considered. Section 7.3 develops a new approach for cost efficiency measurement that considers that both input and price adjustments are at the DMs' discretion. Section 7.4 illustrates the application of the methods developed within the context of the case study of the Portuguese bank branch network. Section 7.5 summarises and concludes.

7.2 Cost efficiency with fixed prices

7.2.1 Prices known at the DMU level

Looking beyond technical efficiency, Farrell (1957) was the first to propose the measurement of cost efficiency by taking into account the economic context in which the DMUs' activity occurs. Farrell (1957) efficiency concepts and the associated DEA models were described in detail in Chapter 2. This section briefly reviews the measurement of cost efficiency based on Farrell (1957) concept. This method is the reference point for the developments described in this chapter.

Based on the example data set reported in Table 7.1, CE measurement can be illustrated using Figure 7.2.

Table 7.1 – *Inputs, output and prices at the DMUs*

	X_1	X_2	Y	P_1	P_2
DMU A	2	7	1	3	4
DMU B	3	5	1	4	3
DMU C	4	4	1	3	4
DMU D	5	3	1	4	3
DMU E	7	2	1	3	4
DMU F	3	7	1	3	4
DMU G	5	5	1	4	3
DMU H	6	3.6	1	4	3
DMU I	9	2	1	4	3
DMU J	10	2.5	1	3	4

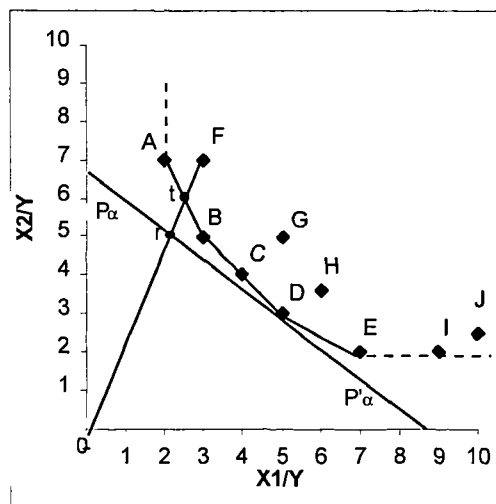


Figure 7.2 – Cost efficiency measurement

Consider DMU F with input prices $P_1=3$ and $P_2=4$. The slope of the isocost $P_\alpha P'_\alpha$ (i.e., $-P_1/P_2$) gives the relative prices at F. The point where this isocost is tangent to the production frontier (point D) identifies the input combination corresponding to the minimal cost for these prices. The CE measure for DMU F is obtained as the ratio of the minimal cost (associated with point D on the frontier of the PPS) to the observed cost (corresponding to the point F within the PPS)¹. Graphically the CE measure is given by the ratio Or/OF , where r has the same cost as D.

The CE measure can be obtained by including weight restrictions in a standard DEA model, as shown in model (7.1)².

¹ The cost at point D and point F is determined for the current prices at F, i.e., $P_1=3$ and $P_2=4$.

² This model can be used as an alternative to Fare *et al.* (1985) model for the estimation of minimal cost. See Chapter 2 for further details on cost efficiency measurement.

Cost efficiency model (with prices fixed and known at each DMU)	(7.1)
$\text{Max } \sum_{r=1}^s u_r y_{rj_0}$ $\text{s.t. } \sum_{i=1}^m v_i x_{ij_0} = 1$ $\sum_{r=1}^s u_r y_{rj} - \sum_{i=1}^m v_i x_{ij} \leq 0, \quad j = 1, \dots, n$ $v_{i^a} - \frac{p_{i^a j_0}}{p_{i^b j_0}} v_{i^b} = 0, \quad i^a < i^b, \quad i^a, i^b = 1, \dots, m \quad (7.1a)$ $u_r \geq \varepsilon, \quad r = 1, \dots, s$	

In the model above, $P_{i^a j_0}$ and $P_{i^b j_0}$ correspond to the prices of input i^a and input i^b , respectively, observed at the DMU j_0 under assessment. Note that restriction (7.1a) imposes that the relative weights underlying the assessment must be identical to the relative input prices observed at the DMU j_0 under assessment (i.e., $\frac{v_{i^a}}{v_{i^b}} = \frac{p_{i^a j_0}}{p_{i^b j_0}}$).

The Farrell CE measure assumes that the prices are fixed and known at each DMU, although they can vary from DMU to DMU. This requires the knowledge of exact prices at each DMU, which may be difficult to obtain in actual applications. Another characteristic of Farrell CE is that only the relative values of input prices are relevant for the assessment. An equiproportional increase (or decrease) in the magnitude of input prices has no effect on the CE measure. E.g. in terms of the graphical illustration of DMU F's CE in Figure 7.2, if the input prices increased (or decreased) proportionally, the isocost line $P_\alpha P'_\alpha$ would remain unchanged, and so would the CE measure. Despite these limitations, only a few extensions to Farrell CE measure have appeared in the

literature since its original development. Many of these extensions were presented in the course of a particular study (e.g., Thompson *et al.*, 1996; Schaffnit *et al.*, 1997) and have been proposed as a solution to overcome inconsistencies or incompleteness of the data.

7.2.2 Prices unknown at the DMU level

Recent research, particularly in the banking sector, enhanced CE models to overcome incompleteness of data, where only the maximal and minimal bounds of market prices were known, see Thompson *et al.* (1996), Thompson *et al.* (1997), Taylor *et al.* (1997) and Schaffnit *et al.* (1997). In these cases, the lack of information on individual DMUs input prices has been effectively addressed by incorporating in the DEA models the price information available as weights restrictions, in the form of “input cone assurance regions” (first developed by Thompson *et al.* (1986) and defined more precisely in Thompson *et al.* (1990)). This enabled a reorientation of the technical efficiency assessment towards cost efficiency measurement. Based on this approach, CE is assessed in the light of the most favourable price scenario (“optimistic” perspective) for the DMUs’ current input mix, provided that the prices (weights) underlying the assessment are within the relative bounds specified. This measure, described in more detail in the following sections, will be referred to as Optimistic Cost Efficiency in the remainder of this thesis.

The aim of this section is to enhance CE measurement in situations of price uncertainty by providing an estimate of CE in the light of the least favourable price scenario for the DMUs’ current input mix (“pessimistic” perspective). This measure will be called Pessimistic Cost Efficiency in the remainder of this thesis.

In situations of price uncertainty, if the range of input prices underlying the CE assessment contains the actual prices at the DMUs, the resulting Optimistic and Pessimistic CE measures can be regarded as the maximal and minimal bounds of a confidence interval containing Farrell CE measure.

The next section provides a pictorial illustration of efficiency measurement in the ‘weights’ (or price) space. This introduction is important to ground the formulation of the DEA models for measuring Optimistic CE and Pessimistic CE described in the following sections.

7.2.2.1 Graphical illustration of efficiency measurement in the ‘weights’ space

This section provides an interpretation of technical and cost efficiency measures (Farrell CE, Optimistic CE and Pessimistic CE) in the ‘weights’ (or price³) space. Instead of representing the efficiency measure in the ‘input’ space, which enables interpreting the ‘envelopment’ formulation of the DEA model, we represent the efficiency measure in the ‘weights’ space, associated with the interpretation of the ‘weights’ formulation of the DEA model.

Since under CRS the DEA input and output oriented models give a similar efficiency score, we have chosen to represent an output-oriented model. This enables representing an efficiency assessment with normalised output weights, providing a clearer interpretation of the input weights.

³ Recall that the ‘weights’ of a DEA assessment can be interpreted as normalised shadow prices.

Figure 7.3 shows the pictorial representation (in the ‘weights’ space) of the DMUs from Table 7.1. The calculations leading to this pictorial representation are detailed in the Appendix 7A. For illustrative purposes, the normalisation constant used for this representation is equal to the minimal cost of output production observed in the DMUs under analysis⁴ (i.e., the value used was 27, which corresponds to the cost observed at DMU B). This ensures that the weights underlying the assessment have a magnitude that is meaningful under the market conditions where the DMUs operate.

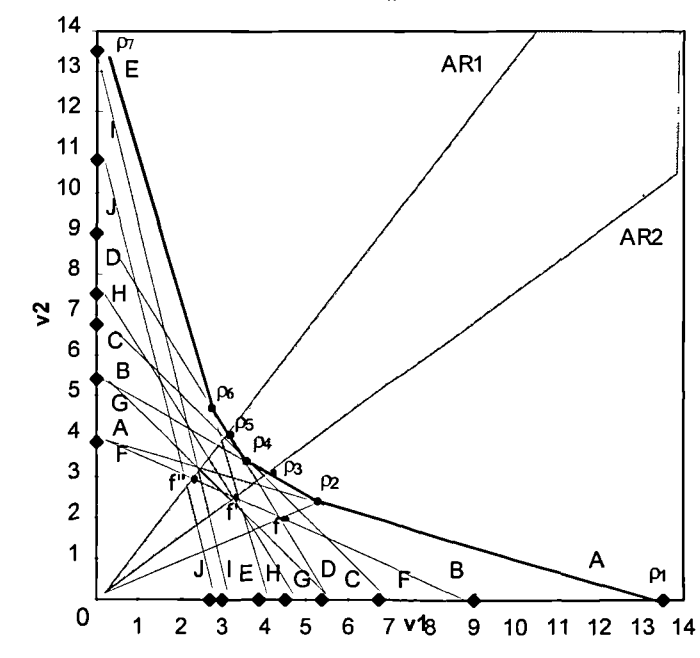


Figure 7.3 - Price (weights) space for the DEA efficiency assessment

Each DMU is represented in Figure 7.3 by a line, corresponding to the linear constraints in the DEA ‘weights’ model. The slope of the line depends on the relative magnitudes of the two inputs used at each DMU, and the distance from the origin depends on their

⁴ The normalisation constant generally used in the DEA models is equal to one. See for example model (2.8) in Chapter 2, with $\sum_{r=1}^s u_r y_{rj_0} = 1$. Using a normalisation constant equal to one has the advantage of obtaining directly the efficiency measure expressed as a percentage.

absolute values (i.e., these lines will be further away from the origin for smaller input values).

The technical efficiency frontier is represented with a bold line in Figure 7.3. It is determined by the DMUs furthest away from the origin in the ‘weights’ space (e.g. DMUs A, B, C, D and E) and corresponds to the piecewise segment defined by $\rho_1, \rho_2, \rho_3, \rho_4, \rho_5, \rho_6$ and ρ_7 . Note that the feasible region of the DEA model defined corresponds to the area to the right and above of this frontier. The DEA model solved for each DMU searches for the minimisation of an objective function whose slope is identical to the constraint associated with that DMU.

Note that the DMUs that are only radially efficient (without however being efficient in Koopmans (1957) sense due to the existence of slacks in input levels), are associated with a constraint in the ‘weights’ space that coincides with the TE frontier only on its intersection with the axes⁵.

The **technical efficiency** in the ‘weights’ space is measured along a ray that goes through the origin. This ray is the one that enables the closest projection of the assessed DMU towards the frontier, such that the evaluation is made under the best possible light. For example, the technical efficiency of DMU F is given by the ratio $\frac{Of}{Op_2}$.

In relation to the weights underlying the TE assessment, the efficient DMUs will generally have multiple optimal solutions. These weights can be directly obtained from

⁵ For DMU I this occurs at point ρ_7 , where the input with a slack value (i.e., input 1) is given zero weight.

Figure 7.3. For example, for DMU A all weight combinations on the segment between ρ_1 and ρ_2 will evaluate it as 100% technical efficient.

Conversely, an inefficient DMU will generally have a unique set of weights that shows it in the best possible light. For example, the optimal set of weights underlying the output-oriented assessment of DMU F is obtained at point ρ_2 . Generalising the interpretation of the weights for an inefficient DMU for a multi-input multi-output situation, (with a normalisation constant equal to minimal cost of production), the optimal weights of an output oriented model represent the prices that would ensure output production at minimal cost with technical efficient input levels⁶.

In relation to the interpretation of **Farrell CE** measure, recall that it assumes that the prices at each DMU are fixed and known, although possibly different among the DMUs. Also, only the relative values of input prices are relevant for the assessment. The interpretation of the CE measure in the ‘weights’ space will be based on the DEA formulation represented in (7.1). Consider again DMU F, with prices $P_1=3$ and $P_2=4$. The relative prices ($P_1/P_2=3/4$) are represented in Figure 7.3 by the ray AR1 (where $v_1/v_2=3/4$). Farrell CE measure for DMU F must be estimated along this ray⁷, and it is equal to $\frac{Of''}{Op_5}$.

In relation to the estimation of **CE with price uncertainty**, consider a scenario where only the maximal and minimal price bounds for all DMUs can be identified, e.g.,

⁶ For an input oriented model using the same normalisation constant, the weights at the optimal solution represent the prices that would ensure output production at minimal cost with the current input levels. For DMU F, the weights of an input oriented assessment are those corresponding to point f .

⁷ As imposed by the restriction (7.1a) in the CE model presented in (7.1).

$P_1^{\min} = P_2^{\min} = 3$ and $P_1^{\max} = P_2^{\max} = 4$. Based on this information, the price ratios underlying the cost efficiency evaluation would be restricted to the following range:

$\left(\frac{P_1^{\min}}{P_2^{\max}} \leq \frac{v_1}{v_2} \leq \frac{P_1^{\max}}{P_2^{\min}} \right)$. For the data considered, this becomes $\left(\frac{3}{4} \leq \frac{v_1}{v_2} \leq \frac{4}{3} \right)$. These

restrictions define an input cone assurance region, represented by the shaded area between the rays AR1 and AR2 in Figure 7.3. An Optimistic CE measure would adopt the most favourable weights combination within this cone to evaluate the DMUs. Conversely, a Pessimistic CE assessment would consider the least favourable price scenario.

In the case of DMU F, the Optimistic CE measure would be given by $\frac{Of'}{Op_3}$, whereas the

Pessimistic CE measure would be equal to $\frac{Of''}{Op_5}$. These correspond to the closest and

furthest radial projections to the efficiency frontier, respectively, within the restricted range of price ratios specified.

The next sections describe the DEA models for measuring Optimistic and Pessimistic CE, as pictorially illustrated above.

7.2.2.2 The optimistic perspective of CE measurement

The model for CE measurement with price uncertainty adopting an optimistic perspective is shown in (7.2). Using this model, CE is assessed in the light of the most favourable price scenario for the input-output mix of each DMU.

Optimistic CE model (with an input orientation)	(7.2)
$\text{Max } \sum_{r=1}^s u_r y_{rj_0}$ $\text{s.t. } \sum_{i=1}^m v_i x_{ij_0} = 1$ $\sum_{r=1}^s u_r y_{rj} - \sum_{i=1}^m v_i x_{ij} \leq 0, \quad j = 1, \dots, n$ $\frac{P_{i^a j_0}^{\min}}{P_{i^b j_0}^{\max}} \leq \frac{v_{i^a}}{v_{i^b}} \leq \frac{P_{i^a j_0}^{\max}}{P_{i^b j_0}^{\min}}, \quad i^a < i^b, \quad i^a, i^b = 1, \dots, m \quad (7.2a)$ $u_r \geq \varepsilon, \quad r = 1, \dots, s \quad (7.2b)$	

The restrictions imposed to the weights underlying the assessment using model (7.2) assume the form of an input cone assurance region (IC-AR). This type of weights restrictions, as defined in (7.2a), was first developed by Thompson *et al.* (1986) and defined more precisely in Thompson *et al.* (1990). An IC-AR is specified as a set of homogeneous inequalities for separable cones, which define an acceptable input value system to underlie the efficiency assessment. This value system is reflected by the relative values of the input weights at the optimal solution to the DEA model. The bounds to the relative input weights imposed by the IC-AR can be based on input price information, moving the focus of the DEA assessment from technical efficiency towards cost efficiency measurement.

In the optimistic CE model defined above, $P_{i^a j_0}^{\min}$ and $P_{i^a j_0}^{\max}$ stand for the minimal and maximal bounds estimated for the price of input i_a at DMU j_0 , respectively. The input cone assurance region defined in (7.2a) imposes that the relative value of the weights underlying the assessment must be within the bounds of the relative prices considered.

The IC-AR specified can be rewritten in linear form, giving the following constraints:

$$\begin{cases} v_{i^a} - \frac{P_{i^a j_0}^{\max}}{P_{i^b j_0}^{\min}} v_{i^b} \leq 0 \\ -v_{i^a} + \frac{P_{i^a j_0}^{\min}}{P_{i^b j_0}^{\max}} v_{i^b} \leq 0 \end{cases} \quad i^a < i^b, \quad i^a, i^b = 1, \dots, m$$

If the DMUs evaluation is based on m inputs, there are C_2^m different ratios between two inputs, which gives a total of $2 \times C_2^m$ linear inequality constraints for defining the IC-AR.

In order to illustrate the differences between Optimistic CE and Farrell CE in relation to efficiency measurement interpreted in terms of the production possibility set, (i.e., the interpretation associated with the ‘envelopment’ (or dual) formulation of model (7.2)), consider Figure 7.4. The DMUs shown are those presented in Table 7.1, whose efficiency measurement in the weights space was illustrated with Figure 7.3.

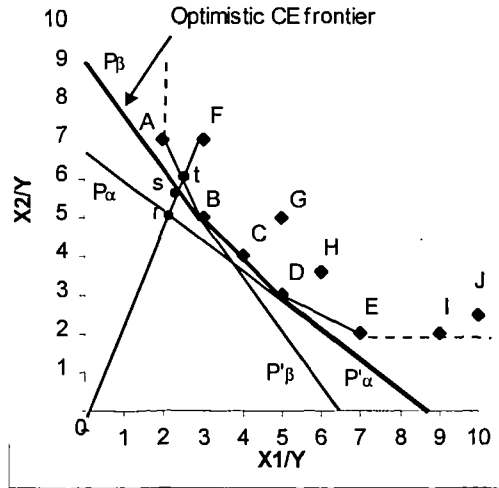


Figure 7.4 – *Optimistic CE versus Farrell CE*

For the calculation of Farrell CE, the individual DMUs input prices must be known. Based on the price information reported in Table 7.1, Farrell CE measure would be

obtained by comparison to segment $P_\alpha P'_\alpha$ (slope $-3/4$) in the case of DMUs A, C, E, F and J, and to segment $P_\beta P'_\beta$ (slope $-4/3$) for DMUs B, D, G, H and I.

Conversely, the Optimistic measure accounts for situations of unknown prices at the DMU level. For illustration purposes, consider again the scenario where the only price information available indicates that the prices of inputs 1 and 2 were between 3 and 4, (i.e., $P_1^{\min} = P_2^{\min} = 3$ and $P_1^{\max} = P_2^{\max} = 4$), without being possible to specify the exact prices at each DMU. In this example, the slope of the isocost underlying the measurement of Optimistic CE could vary between the slope of $P_\beta P'_\beta$ (i.e., $-\frac{P_1^{\max}}{P_2^{\min}} = -\frac{4}{3}$) and the slope of $P_\alpha P'_\alpha$ (i.e., $-\frac{P_1^{\min}}{P_2^{\max}} = -\frac{3}{4}$). The Optimistic CE measure assesses each DMU by comparison to the most favourable isocost line within the bounds specified (i.e., the isocost that has an underlying price ratio as close as possible to the marginal rate of substitution between the inputs). In Figure 7.4, the Optimistic CE frontier would correspond to the segments linking P_β , B, C, D and P'_α . In the case of DMU F, the Optimistic CE measure is given by Os/OF (note that point s is on the isocost associated with the most favourable relative prices for DMU F), whereas Farrell CE measure is equal to Or/OF (note that point r is on the isocost associated with the relative prices observed at DMU F).

7.2.2.3 The pessimistic perspective of CE measurement

The previous section described the evaluation of CE under an “optimistic” perspective. The aim of this section is to extend the CE measurement method in situations of price uncertainty by providing a model for the estimation of the Pessimistic cost efficiency.

This involves adopting for the assessment the least favourable price scenario within the range of input prices considered.

These two estimates of CE, corresponding to the “optimistic” and “pessimistic” perspectives, constitute the maximal and minimal bounds of a confidence interval for the Farrell CE measure, assuming that the actual prices paid at the DMUs are within the relative bounds considered in the DEA assessment with price uncertainty.

In order to obtain the Pessimistic CE measure, it is not enough to change the objective function of the Optimistic CE model (7.2) from a maximisation to a minimisation. Note that this would give an efficiency estimate approximately zero for all DMUs assessed. The only binding constraints preventing the efficiency value to be equal to zero would be (7.2b).

Therefore, the computation of the Pessimistic CE measure requires developing an alternative formulation of the DEA model. This new formulation proposed is described next.

We start developing an alternative formulation for the standard DEA ‘weights’ model for TE measurement. This model is then modified to enable the estimation of CE (i.e., Farrell CE, Optimistic CE and Pessimistic CE).

Recall that the efficiency of a DMU j_0 is a relative measure, obtained by comparison to the “best practices” observed in other peer DMUs in the set under analysis. These peers must have an efficiency score equal to 1, subject to the constraints that all other DMUs in the set have an efficiency value smaller or equal to one when evaluated with the same weights. The objective function of this model searches for the maximal value of the

efficiency score of DMU j_0 when measured against peers satisfying the constraints stated above⁸.

The alternative method proposed in this section for obtaining a DEA efficiency score requires solving more than one linear program. The assessment consists of running a set of linear programming models, corresponding to partial optimisations, where each DMU in the set is considered in turn as a peer.

As it is not possible to identify *a-priori* the suitable peer DMUs in the set under analysis, the assessment of a DMU j_0 involves solving a set of n models⁹ such as (7.3), considering, one at a time, all the DMUs in the set as potential peers for DMU j_0 .

Alternative formulation of the DEA 'weights' model under CRS		(7.3)
$\text{Max } \psi_{j_0 j_p} = \sum_{r=1}^s u_r y_{rj_0}$		
$\text{s.t. } \sum_{i=1}^m v_i x_{ij_0} = 1$		
$\sum_{r=1}^s u_r y_{rj_p} - \sum_{i=1}^m v_i x_{ij_p} = 0$		(7.3a)
$\sum_{r=1}^s u_r y_{rj} - \sum_{i=1}^m v_i x_{ij} \leq 0,$		$j = 1, \dots, n$
$v_i \geq \varepsilon,$		$i = 1, \dots, m$
$u_r \geq \varepsilon,$		$r = 1, \dots, s$
		(7.3b)

⁸ This corresponds to the interpretation of the 'weights' formulation of the DEA model (see model (2.7) in Chapter 2).

⁹ Note that the number of models (n) that needs to be solved for assessing any DMU j_0 is equal to the total number of DMUs in the set under assessment.

The index j_p in model (7.3) represents the peer DMU considered for the estimation of the efficiency of DMU j_0 . The constraint (7.3a) forces the efficiency of the peer DMU j_p to be equal to one in the assessment of DMU j_0 . As a result, model (7.3) may not have a feasible solution, which indicates that DMU j_p is not suitable as a peer¹⁰ for DMU j_0 .

The **technical efficiency** measure is obtained as the maximal score $\psi_{j_0 j_p}^*$ among all optimal solutions to model (7.3), associated with the use of peers that render the model feasible.

This procedure is an alternative to the standard DEA model (Charnes *et al.*, 1978), although in order to obtain an efficiency score for the n DMUs in the set under analysis, it requires solving n^2 linear programming models.

By replacing the restrictions in (7.3b) with the restrictions in (7.1a), the procedure described above provides an estimate of **Farrell CE**. This requires the knowledge of exact prices at each DMU.

Replacing the restrictions in (7.3b) with the input cone assurance region restrictions in (7.2a), the procedure above provides an estimate of the **Optimistic CE**.

Finally, replacing the restrictions in (7.3b) with the input cone assurance region restrictions in (7.2a), and changing the objective function of (7.3) from a maximisation to a minimisation, the Pessimistic CE measure is obtained choosing the minimal score

¹⁰ Model (7.3) will only be feasible if the DMU j_p used as peer is located on the frontier of the PPS.

$\psi_{j_0j_p}$ among all peers for which the model is feasible. This Pessimistic CE model is shown in (7.4).

Pessimistic CE model	(7.4)
$\text{Min } \psi_{j_0j_p} = \sum_{r=1}^s u_r y_{rj_0}$ $\text{s.t. } \sum_{i=1}^m v_i x_{ij_0} = 1$ $\sum_{r=1}^s u_r y_{rj_p} - \sum_{i=1}^m v_i x_{ij_p} = 0$ $\sum_{r=1}^s u_r y_{rj} - \sum_{i=1}^m v_i x_{ij} \leq 0, \quad j = 1, \dots, n$ $\frac{P_{i^a j_0}^{\min}}{P_{i^b j_0}^{\max}} \leq \frac{v_{i^a}}{v_{i^b}} \leq \frac{P_{i^a j_0}^{\max}}{P_{i^b j_0}^{\min}}, \quad i^a < i^b, \quad i^a, i^b = 1, \dots, m$ $u_r \geq \varepsilon, \quad r = 1, \dots, s$	

This model is valid in the special case of an input homothetic technology¹¹. If the technology is not homothetic the cost-minimising input proportions depend on output levels (see Schmidt, 1985). As the pessimistic CE measure results from an assessment that shows the DMUs in the least favourable light, with multiple output dimensions the input levels against which the DMUs can be compared can lead to very low Pessimistic CE measures.

In real world assessments based on multiple inputs and outputs, the technology is generally not homothetic¹². Consequently, the Pessimistic CE measure may result from an efficiency evaluation with respect to points on the PPS that show the DMU under

¹¹ A technology is input homothetic iff each isoquant is a radial expansion of the unit isoquant.

¹² This property can be tested using econometrics.

assessment in a very unfavourable light. This can lead to very low Pessimistic CE estimates, without a clear managerial interpretation.

In CRS assessments involving multiple inputs and a single output the technology is always homothetic. Thus, in real world applications, the Pessimistic CE model should only be used for assessments involving a single output.

The interpretation of the Pessimistic CE measure (in the input space, corresponding to a representation of the production possibility set) is illustrated in Figure 7.5.

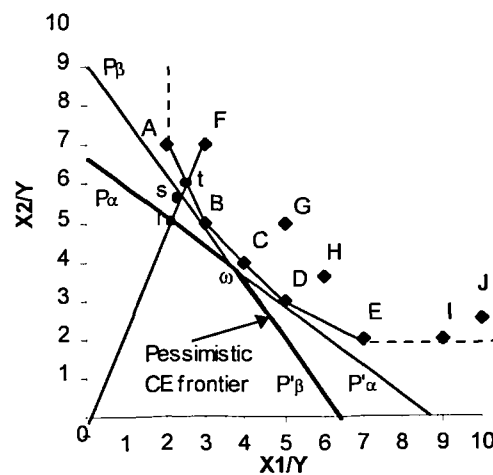


Figure 7.5 – Illustration of Pessimistic CE

Recall that considering the range of input prices as defined in the previous section, the Optimistic CE measure is defined by the segments linking P_β , B, C, D and P'_α . Conversely the Pessimistic CE measure assesses each DMU by comparison to the least favourable price scenario. Thus, the Pessimistic CE frontier is defined by the segments linking P_α , ω and P'_β . In the case of DMU F, the Optimistic CE measure is given by O_s/OF , whereas both Farrell CE measure and the Pessimistic measure would be equal to O_r/OF . Note that the Pessimistic and Optimistic CE measures define the bounds of a confidence interval that contains the Farrell CE measure.

7.3 Cost efficiency with variable prices

7.3.1 Introduction

The CE measures described in the previous section assume that prices are fixed. If the prices are known at the DMU level, it is possible to obtain a precise estimate of CE (i.e., Farrell CE measure). If there is price uncertainty the CE assessment involves the estimation of a confidence interval, whose minimal and maximal bounds are equal to the Pessimistic and Optimistic CE measures, respectively.

All the CE measures discussed previously (i.e., Farrell CE, Optimistic CE and Pessimistic CE) only reflect input efficiency, i.e. the extent to which the input levels enable production with minimal cost under certain price conditions. The inefficiencies detected are either due to using too much input (technical inefficiency) or due to having the wrong mix of inputs in the light of prices (input allocative inefficiency).

The aim of this section is to enhance CE assessment to address situations where both input and price adjustments are at the DMs' discretion. Although controllability of input prices is an assumption rarely adopted in economics, in the context of the bank branches' operational activity having some control over prices is a realistic assumption.

Considering the input set defined under the 'production' approach, representing the number of employees at each branch, salary adjustments are possible on filling new posts and through negotiation with employees. By focusing on changes to input prices instead of input levels, this approach to CE assessment can suggest salary reductions or freezes rather than firing.

At the organisational level, the effort to keep current employees by focusing on the negotiation of employment conditions may be beneficial, as it avoids the difficulties associated with discharging or reallocating staff within the bank. At the DMU level, less radical changes to the input mix enable, to some extent, the maintenance of current practices.

In a context of variable prices, a new economic efficiency measure is proposed. This measure reflects both price efficiency and input efficiency. The next section illustrates these measures and describes the DEA models required for their calculation.

A target setting method for attaining economic efficiency through adjustments to input and price levels is described in the final section (Section 7.3.3).

7.3.2 The economic efficiency components

The measurement of economic efficiency under variable prices is pictorially illustrated using the data set in Table 7.2. The production possibility set defined by these DMUs is shown in Figure 7.6.

Table 7.2 – *Example data set*

	X_1	X_2	Y	P_1	P_2
DMU A	2	7	1	9	6.2
DMU B	3	5	1	7	9
DMU C	5	3	1	9	6.2
DMU D	7	2	1	4	14
DMU E	9	2	1	7	9
DMU F	3	7	1	7	9
DMU G	5	5	1	9	6.2

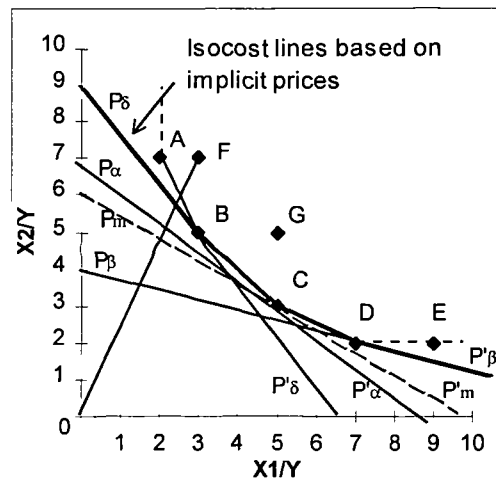


Figure 7.6 – Production possibility set for the example considered

The efficient frontier is represented by the segments defined by points A, B, C and D.

$P_\alpha P'_\alpha$ represents the isocost for an evaluation of CE with the prices observed at DMUs B, E and F (i.e., $P_1 = 7$ and $P_2 = 9$). $P_\beta P'_\beta$ is the isocost associated with the prices observed at DMU D (i.e., $P_1 = 4$ and $P_2 = 14$) and $P_\delta P'_\delta$ is the isocost associated with the prices observed at DMUs A, C and G (i.e., $P_1 = 9$ and $P_2 = 6.2$).

Considering that all DMUs are operating under the same market conditions, the DMUs paying prices above the minimal levels observed (i.e., $P_1^{\min} = 4$ and $P_2^{\min} = 6.2$) cannot be considered fully efficient. Paying higher prices than those observed elsewhere in the market represents inefficiency in prices, indicating that the DMUs are slow to adjust their input prices to more favourable market conditions.

The notion of economic efficiency and its decomposition under variable prices is introduced with an illustrative example using DMU F. Its current cost is equal to 84. The

minimal cost achievable with its current prices is equal to 62¹³. Finally, the minimal cost achievable with the minimal prices observed in other DMUs that operate in the same market is equal to 38.6¹⁴. Economy efficiency represents the proportional reduction to total cost achievable with changes to input levels and prices. This measure is equal to 46% (38.6/84). It can be decomposed into input efficiency, which is equal to Farrell CE measure, and price efficiency, representing the proportional reduction to cost achievable from price adjustments. Input efficiency is equal to 74% (62/84) and price efficiency is equal to 62% (38.6/62). The product of price efficiency and input efficiency is equal to the economic efficiency measure.

The economic efficiency measure under variable prices and its components of price efficiency and input efficiency can be defined more precisely as follows:

- Economic efficiency measures the ability to produce current outputs at minimal cost, evaluated with the minimal input prices observed under the current market conditions. It is obtained as follows:

$$\text{Economic efficiency} = \frac{\text{Minimal cost with minimal prices}}{\text{Observed cost}}. \quad (7.5)$$

- Input efficiency reflects the ability to produce at the minimal cost, given the current price levels at each DMU. It is equal to the Farrell CE measure, obtained as follows:

$$\text{Input efficiency} = \frac{\text{Minimal cost with current prices}}{\text{Observed cost}}. \quad (7.6)$$

¹³ Note that the isocost line $P_{\alpha}P'_{\alpha}$, based on the prices observed at DMU F (i.e., $P_1 = 7$ and $P_2 = 9$) is tangent to the PPS at point C. The cost at point C, evaluated with these prices, is equal to 62.

¹⁴ Note that the isocost line $P_mP'_m$, based on the minimal prices observed in the data set under analysis (i.e., $P_1 = 4$ and $P_2 = 6.2$) is tangent to the PPS at point C, whose cost of production, evaluated with these prices, is equal to 38.6.

Note that the minimal cost with current prices is obtained at the optimal solution to model (2.20), as described in Chapter 2.

- Price efficiency captures the extent to which the DMUs have adequate input price levels under the current market conditions. Inefficiencies associated with overpayment of resources can be captured by this measure. It is obtained as follows:

$$\text{Price efficiency} = \frac{\text{Minimal cost with minimal prices}}{\text{Minimal cost with current prices}}. \quad (7.7)$$

In the remainder of this thesis, the minimal cost with minimal market prices will be referred to as the Target Budget of a DMU. It is equal to the cost of producing the current output levels with the minimal input prices associated with the market conditions where the DMU operates¹⁵.

The Target Budget of DMU j_0 is obtained solving model (7.8), where p_i^{\min} represents the minimal price of input i under the market conditions where DMU j_0 operates. The difference between the Fare *et al.* (1985) model for CE measurement described in Chapter 2 (model (2.20)) and model (7.8) is that the latter computes the minimal cost based on minimal market prices instead of the prices currently paid at each DMU.

¹⁵ For example, the DMUs in the small illustrative example in Table 7.2 have a Target Budget equal to 38.6.

Model for deriving the Target Budget		(7.8)
$TB_{j_0} = \text{Min} \sum_{i=1}^m p_i^{\min} x_i^0$		
s.t. $\sum_{j=1}^n x_{ij} \lambda_j = x_i^0,$	$i = 1, \dots, m$	
$\sum_{j=1}^n y_{rj} \lambda_j \geq y_{rj_0},$	$r = 1, \dots, s$	
$\lambda_j \geq 0,$	$j = 1, \dots, n$	
$x_i^0 \geq 0,$	$i = 1, \dots, m$	

Note that in a DEA study involving a large number of DMUs, they may be operating in a few markets with different characteristics. In this case, the minimal input prices (p_i^{\min}) considered in the assessment of DMU j_0 must have been observed in other DMUs operating under similar market conditions.

The product of input efficiency and price efficiency components gives the economic efficiency measure in a scenario of variable prices. The decomposition of economic efficiency is illustrated in Figure 7.7

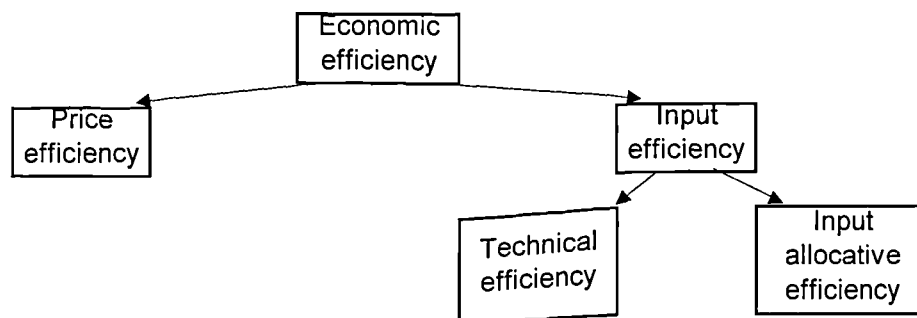


Figure 7.7 – Components of economic efficiency under a variable price scenario

7.3.3 Target setting with emphasis on keeping the current input mix

This section focuses on the implementation of economic efficient targets such that production at minimal cost (i.e., the Target Budget of each DMU) can be achieved causing as little disturbance as possible to the current practices at the DMUs.

The targets proposed in this section are an alternative to the targets obtained as by-products from model (7.8). Recall that using model (7.8) the input targets for DMU j_0 are equal to x_i^{0*} , $i=1,\dots,m$, and the price targets are equal to p_i^{\min} , $i=1,\dots,m$. This combination of input and price levels enables the achievement of the Target Budget (TB $_{j_0}$).

The method developed in this section favours the modification of price levels in exchange for keeping the input mix as close as possible to the current mix. These price targets must satisfy the following properties:

- The relative prices must be as close as possible to the marginal rate of substitution between the inputs.
- The relative prices must have been observed in the market.
- The target price levels must ensure the attainment of economic efficiency, i.e. production with the Target Budget.

The target price levels with the above properties are obtained solving model (7.9). These will be referred to as 'implicit prices' in the remainder of this thesis:

<p style="text-align: right;">Model for the identification of ‘implicit prices’ leading to economic efficiency (7.9)</p> $\text{Min } \sum_{i=1}^m v_i x_{ij_0}$ $\text{s.t. } \sum_{r=1}^s u_r y_{rj_0} = TB_{j_0}$ $\sum_{r=1}^s u_r y_{rj} - \sum_{i=1}^m v_i x_{ij} \leq 0, \quad j = 1, \dots, n$ $\frac{p_{i^a}^{\min}}{p_{i^b}^{\max}} \leq \frac{v_{i^a}}{v_{i^b}} \leq \frac{p_{i^a}^{\max}}{p_{i^b}^{\min}}, \quad i^a < i^b, \quad i^a, i^b = 1, \dots, m \quad (7.9a)$ $u_r \geq \varepsilon, \quad r = 1, \dots, s$

The ‘implicit prices’ of DMU j_0 under assessment are equal to the input weights at the optimal solution (i.e. v_i^* , $i = 1, \dots, m$). The normalising constant used (TB_{j_0}) is equal to the Target Budget, obtained at the optimal solution to model (7.8).

Note that model (7.9) is the output oriented version of the Optimistic CE model¹⁶. This implies that the weights obtained at the optimal solution are as close as possible to the marginal rate of substitution between the inputs, as the evaluation of Optimistic CE shows the DMUs in the best possible light.

The IC-AR (7.9a) imposes bounds to the relative values of the ‘implicit prices’, in order to ensure that they are feasible under the current market conditions.

The output orientation of model (7.9), associated with the use of a normalisation constant equal to the Target Budget, ensures that the weights obtained at the optimal solution are

¹⁶ See section 7.2.2.2 for details regarding Optimistic CE.

equal to the prices that enable having an input cost equal to the Target Budget (TB_{j_0}), if operating with technical efficient input levels¹⁷.

So far we have concentrated on the identification of the ‘implicit prices’. In order to complete the target setting method, it is also important to identify the input levels that allow production with the Target Budget, in the light of these prices.

The target input levels are obtained from the dual formulation (multiplier form) of an input oriented Optimistic CE model, such as (7.2). Model (7.2) is reproduced below, with the input-cone assurance region represented in a condensed form. Note that the restriction in (7.10a) is identical to the one in (7.9a).

Optimistic CE model		(7.10)
$\text{Max } \sum_{r=1}^s u_r y_{rj_0}$		
$\text{s.t. } \sum_{i=1}^m v_i x_{ij_0} = 1$		
$\sum_{r=1}^s u_r y_{rj} - \sum_{i=1}^m v_i x_{ij} \leq 0,$		$j = 1, \dots, n$
$\sum_{i=1}^m z_{ci} v_i \leq 0,$		$c = 1, \dots, k$
$u_r \geq \varepsilon,$		$r = 1, \dots, s$

The parameters z_{ci} result from a condensed representation of the IC-AR in matrix format, where c is the line index and i is the column index, corresponding to each of the

¹⁷ See section 7.2.2.1 for further details regarding the interpretation of the weights in an output-oriented DEA model.

linear constraints of the IC-AR. k represents the number of constraints needed to represent the IC-AR ($k = 2 \times C_2^m$).

The dual to model (7.10) used for the identification of the input targets associated with the ‘implicit prices’ is shown in (7.11).

<p style="text-align: right;">Model for the identification of the input targets associated with the ‘implicit prices’ (7.11)</p> $\begin{aligned} &\text{Min } \theta - \varepsilon \sum_{r=1}^t s_r \\ &\text{s.t. } \theta x_{ij_0} - \sum_{j=1}^n \lambda_j x_{ij} + \sum_{c=1}^k z_{ci} \tau_c = 0, \quad i = 1, \dots, m \\ &\quad \sum_{j=1}^n \lambda_j y_{rj} - s_r = y_{rj_0}, \quad r = 1, \dots, t \\ &\quad \lambda_j, \tau_c, s_r \geq 0, \quad \forall j, c \text{ and } r \end{aligned}$
--

The variables in the model above are λ_j, τ_c, s_r and θ . For DMU j_0 , the input targets located on the frontier of the production possibility set are obtained using the following formula:

$$x_{ij_0}^{\text{TIP}} = \theta x_{ij_0} + \sum_{c=1}^k z_{ci} \tau_c = \sum_{j=1}^n \lambda_j x_{ij}, \quad i = 1, \dots, m \quad (7.12)$$

The target setting method described before is illustrated with a small example in the next section.

7.3.3.1 Illustration of the target setting method

In order to illustrate the target setting method described in the previous section, leading to production with a cost equal to the Target Budget involving minimal changes to the

current input mix of the DMUs, consider the data set in Table 7.2 and represented in Figure 7.6.

Recall that in this example, the Target Budget (TB_{j_0}) is equal to 38.6 for all DMUs.

The estimation of the ‘implicit prices’ and associated input targets require the definition of acceptable bounds for the relative prices, which are included in the DEA models (7.9) and (7.11) in the form of an input cone assurance region. Based on the example considered, the input-cone assurance region would be defined as follows¹⁸:

$$\frac{4}{14} \leq \frac{v_1}{v_2} \leq \frac{9}{6.2}, \text{ or in linear form } \begin{cases} v_1 - \frac{9}{6.2} v_2 \leq 0 \\ -v_1 + \frac{4}{14} v_2 \leq 0 \end{cases}$$

The representation of this input cone assurance region in condensed form, as in (7.10a), involves the use of the following parameters:

$$z_{11} = 1, z_{12} = -\frac{9}{6.2}, z_{21} = -1, z_{22} = \frac{4}{14}$$

In terms of the graphical interpretation of the ‘implicit prices’ and input targets, recall that the relative value of the ‘implicit prices’ is as close as possible to the marginal rate of substitution for the inputs of the DMU under analysis. The slope of the segments in bold in Figure 7.6 represents the relative values of the ‘implicit prices’ for the different input mixes of the DMUs in this production possibility set.

¹⁸ The bounds of the input cone assurance region defined in (7.9a) and (7.11a) would be based on the maximal and minimal prices observed for inputs 1 and 2, e.g., $\left(\frac{P_1^{\min}}{P_2^{\max}} \leq \frac{v_1}{v_2} \leq \frac{P_1^{\max}}{P_2^{\min}} \right)$.

Note that the relative values of the input prices are within the values observed in the data set considered (i.e., the slope of the segments in bold is between the slopes of isocosts $P_\delta P'_\delta$ and $P_\beta P'_\beta$).

The target input levels, obtained from model (7.10), correspond to the point where the isocost with the most favourable price ratio for each DMU is tangent to the efficiency frontier. For the DMUs considered, the target input levels are located on the segments linking B, C and D (i.e., the part of the PPS frontier in bold in Figure 7.6).

Table 7.3 shows the economic efficient targets (with ‘implicit prices’) obtained with the above procedure. The alternative economic efficient targets (with the minimal market prices) obtained from the optimal solution to model (7.8) are also shown.

Table 7.3 - Targets with variable prices

DMUs	X_1	X_2	Minimal P_1	Minimal P_2	Cost	X_1	X_2	‘implicit’ P_1	‘implicit’ P_2	Cost
A	5	3	4	6.2	38.6	3	5	6.0	4.1	38.6
B	5	3	4	6.2	38.6	3	5	6.0	4.1	38.6
C	5	3	4	6.2	38.6	5	3	4.8	4.8	38.6
D	5	3	4	6.2	38.6	7	2	3.5	7.0	38.6
E	5	3	4	6.2	38.6	7	2	2.8	9.7	38.6
F	5	3	4	6.2	38.6	3	5	6.0	4.1	38.6
G	5	3	4	6.2	38.6	4	4	4.8	4.8	38.6

Farrell CE targets are shown in Table 7.4, to enable the comparison of the two approaches based on fixed and variable prices assumptions.

Table 7.4 - Targets with fixed prices

DMUs	X_1	X_2	P_1	P_2	Cost
A	3	5	9	6.2	58
B	5	3	7	9	62
C	3	5	9	6.2	58
D	7	2	4	14	56
E	5	3	7	9	62
F	5	3	7	9	62
G	3	5	9	6.2	58

The case of DMU F will be explored in greater detail. As reported in Table 7.3, one way to achieve the Target Budget is to reduce the prices to the minimal values observed in the network ($P_1=4$ and $P_2 = 6.2$) and operate at point C on the frontier of the PPS. These targets would be obtained solving model (7.8). However, these targets involve a substantial change to the input mix currently used at DMU F. The adjustment of relative prices would represent moving from the isocost $P_\alpha P'_\alpha$ to the isocost $P_m P'_m$.

The target setting method based on ‘implicit prices’ finds an alternative set of targets that still enables the achievement of the Target Budget whilst keeping the input mix as close as possible to the original resources balance. In this case, the price targets to be adopted by DMU F are the ‘implicit prices’ ($P_1=6.0$ and $P_2 = 4.1$). Note that their relative value is as close as possible to the input marginal rate of substitution associated with DMU F, and equal to the slope of line $P_\delta P'_\delta$ (where the underlying relative prices were observed in the market). The associated input targets on the frontier of the PPS correspond to point B, where the isocost line defined by $P_\delta P'_\delta$ is tangent to the TE frontier. The major limitation of these alternative targets is that they propose paying input 2 below the minimal price observed in the network, although this enables paying input 1 a higher price than the minimal value observed in other DMUs. This is necessary in order to have target input levels as close as possible to the TE input values and still produce with a total cost equal to the Target Budget. If targets below the minimal levels observed are felt to be unattainable, then there will have to be a greater variation to the input levels, as suggested by the targets based on the minimal market prices.

7.4 Illustration of the methods developed with the assessment of bank branches

This section illustrates the efficiency measurement and target setting methods developed in this chapter in the analysis of bank branches' operational activity. The inputs and outputs of the DEA analysis were defined based on the production approach, as described in Chapter 5. The variables used are reproduced below:

- Inputs :**
- Number of branch and account managers.
 - Number of administrative and commercial staff.
 - Number of tellers.
 - Operational costs (excluding staff costs).

- Outputs:**
- Total value of deposits.
 - Total value of loans.
 - Total value of off balance sheet business.
 - Number of general service transactions.

In the case of this network, the data available was very detailed and the input prices were known at the DMU level. The data collected referred to the average annual salaries of staff in each of the groups defined, as follows:

- Input prices:**
- Average salary and fringe benefits of branch and account managers.
 - Average salary and fringe benefits of administrative/commercial staff.
 - Average salary and fringe benefits of tellers.

The input prices were analysed in detail prior to the efficiency assessment. The analysis separated the branches in 4 different regions within Portugal, according to the banks' internal structure of network management. Table 7.5 shows the summary statistics relating to the input prices (in million escudos) for the labour groups defined.

Table 7.5 - *Branches' input prices in different regions of Portugal*

<i>Regions</i>	<i>Employees group</i>	<i>Mean</i>	<i>SD</i>	<i>Min</i>	<i>Max</i>
Lisbon	Branch/account managers	5.33	0.39	4.56	6.21
	Administrative/commercial staff	3.08	0.44	1.54	3.82
	Tellers	4.37	0.29	3.78	5.16
Porto	Branch/account managers	5.27	0.43	4.44	6.04
	Administrative/commercial staff	3.55	0.57	2.30	4.34
	Tellers	4.34	0.28	4.00	4.97
Centre-south	Branch/account managers	5.39	0.50	4.50	6.64
	Administrative/commercial staff	3.41	0.30	2.74	4.04
	Tellers	4.11	0.37	3.08	4.80
North	Branch/account managers	5.49	0.85	3.89	8.32
	Administrative/commercial staff	3.52	0.50	2.23	4.39
	Tellers	4.10	0.50	2.28	4.69
All	Branch/account managers	5.38	0.57	3.89	8.32
	Administrative/commercial staff	3.37	0.47	1.54	4.39
	Tellers	4.22	0.39	2.28	5.16

Table 7.5 suggests that there are some differences in regional prices, as well as a large spread of prices within each region.

The approach adopted for the CE assessment depends on the DM's degree of control over prices and the time scale considered for the implementation of the results. In the short-term, the input prices of these branches should be considered fixed, as salary adjustments involve a negotiation process that is generally long. Thus, in the short-term, the performance assessment and improvement effort should follow an approach considering fixed prices.

If a longer time period is considered, than salary adjustments become feasible and the performance measurement method should allow for changes to both input and price levels.

The remainder of this section reports the CE assessments considering both scenarios of fixed and variable prices, corresponding to a short-term and a long-term period for the implementation of the results, respectively.

7.4.1 Cost efficiency with fixed prices

7.4.1.1 Prices known at the DMU level

The cost efficiency assessment reported in this section follows Farrell (1957) concept, and each branch is evaluated based on current prices. The results discussed earlier in Chapter 5 are reproduced here (see Table 7.6), in order to enable a comparison with the new methods of cost efficiency measurement developed in this chapter

Table 7.6 – *Farrell CE measure and its components*

	<i>Mean</i>	<i>No. efficient branches</i>
Farrell cost efficiency	69%	12
Input allocative efficiency	85%	12
Technical efficiency	81%	41

Table 7.6 shows that the average input allocative efficiency (85%) is higher than technical efficiency (81%), which indicates that the major inefficiencies are due to operating away from the frontier of the production possibility set (i.e., technical inefficiency).

Table 7.7 shows the overall network targets to attain Farrell CE.

Table 7.7 – *Targets to achieve cost efficiency (with current prices)*

<i>Inputs</i>	<i>Initial values</i>	<i>CE targets (% initial value)</i>	
No. branch/account managers	505	341	(68%)
No. administrative/commercial staff	477	386	(81%)
No. tellers	335	291	(87%)
Operational costs	2600	1565	(60%)
<i>Total input cost</i>	<i>8301</i>	<i>5912</i>	<i>(71%)</i>

As shown in Table 7.7 in order to achieve cost efficiency, the number of administrative/commercial staff and tellers must be reduced to approximately 81% and 87% of their initial value, respectively. However, the reductions required to the number of managers and other operational costs are more significant, with targets of approximately 68% and 60% of current levels, respectively. The elimination of technical and allocative inefficiencies would lead to the reduction of the operational costs to 71% of the current level.

The next section illustrates cost efficiency measurement in a scenario of price uncertainty.

7.4.1.2 Prices unknown at the DMU level

This section illustrates CE measurement considering that the exact prices paid at each branch could not be obtained, and only the maximal and minimal regional input prices were available.

For the reasons described earlier in this chapter, the Pessimistic CE model should be used only for assessments involving a single output. Therefore, this section illustrates the computation of the confidence interval for the CE measure based on a reduced model of branches' activity. The output used represents the branches' total volume of business, and is equal to the sum of the value of deposits, loans and off balance sheet business. Thus, the input-output set used for illustrative purposes is as follows:

- Inputs :***
- Number of branch and account managers.
 - Number of administrative and commercial staff.
 - Number of tellers.
 - Operational costs (excluding staff costs).

Output: • Total value of branches' business (including deposits, loans and off balance sheet business).

Input prices: • Average salary and fringe benefits of branch and account managers.
• Average salary and fringe benefits of administrative/commercial staff.
• Average salary and fringe benefits of tellers.

Table 7.8 reports the average Optimistic CE and Pessimistic CE of all branches analysed. The Farrell CE measure obtained from an evaluation with the actual prices at each DMU is also reported, to explore the effect of price uncertainty on the efficiency estimate.

Table 7.8 – *Cost efficiency measures with price uncertainty*

	Mean	No. efficient branches
Optimistic CE	35.4%	1
Farrell CE	33.8%	1
Pessimistic CE	32.4%	1

The average amplitude of the confidence interval defined by the Optimistic and Pessimistic CE measures at each DMU is 3%. Figure 7.8 shows the distribution of the difference between the Optimistic CE and Pessimistic CE measures for the branches analysed. It shows that the uncertainty in input prices is unlikely to generate cost efficiency estimates that differ from the Farrell CE measure (based on the actual prices paid at the each DMU) by more that 6%.

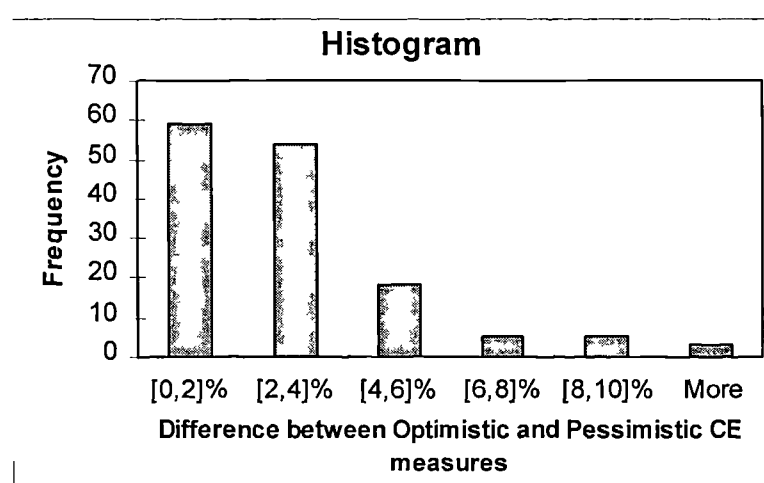


Figure 7.8 – *Difference between the Optimistic CE and Pessimistic CE measures*

7.4.2 Cost efficiency with variable prices

This section reports a cost efficiency assessment considering a longer time period for the implementation of the results, such that salary adjustments are possible. In relation to the targets, changes to both input and price level are possible.

7.4.2.1 Economic efficiency components

The analysis of the distribution of input prices in the four regions considered (see Table 7.5) suggested that they should be treated as different markets. Therefore, the Target Budget of each branch was determined based on the minimal prices observed in its own region. Table 7.9 shows the summary results of economic efficiency and its components for the branch network.

Table 7.9 – *Efficiency measures considering variable prices*

	<i>Mean</i>	<i>No. efficient branches</i>
Economic efficiency	56%	0
Price efficiency	82%	0
Input efficiency	69%	12

The results above indicate that if the branches could pay their inputs at the minimal price observed in other branches in their region, the cost would be, on average, only 82% of the minimal cost level with current prices, and only 56% of the branches' current costs. It can be noted that 100% price efficiency has not been observed in any branch.

The targets associated with the achievement of economic efficiency by paying the minimal prices observed in the branches own region are reported in Table 7.10. The input targets are reported as the total value for the network and the price targets are reported as the average value for all branches analysed.

Table 7.10 - *Targets to achieve economic efficiency with minimal regional prices*

<i>Inputs</i>	<i>Initial values</i>		<i>Target values</i>	
	<i>Total input</i>	<i>Initial prices</i>	<i>Total input</i>	<i>Minimal regional prices</i>
No. branch/account managers	505	5.38	336	4.36
No. administrative/commercial staff	477	3.37	404	2.22
No. tellers	335	4.22	286	3.23
Operational costs	2600	1	1561	1
<i>Total input cost</i>	8301		4818 (58%)	

7.4.2.2 Target setting with emphasis on keeping the current input mix

Within the context of price variability, this section derives the targets that enable output production at the Target Budget, with minimal changes to branches' TE input levels.

The derivation of target prices (i.e., the 'implicit prices') involves solving model (7.9), where the weights restrictions are defined in order to ensure that the relative magnitude of the target prices is within the bounds observed in the branches' own region. The IC-ARs were defined based on the maximal and minimal prices observed in each region. The IC-AR used for the evaluation of branches in Lisbon is described in detail in the appendix 7B.

The target input levels associated with the 'implicit prices' are obtained from model (7.11). Table 7.11 shows the summary results of the target inputs and 'implicit prices'. The input targets are reported as the total value for the network. The price targets are reported as the average value for all branches analysed.

Table 7.11 - *Targets to achieve economic efficiency with minimal changes to input mix*

<i>Inputs</i>	<i>Initial values</i>		<i>Target values</i>	
	<i>Total input</i>	<i>Average price</i>	<i>Total input</i>	<i>Average 'implicit' price</i>
No. branch/account managers	505	5.38	346	4.12
No. administrative/commercial staff	477	3.37	384	2.83
No. tellers	335	4.22	288	3.63
Operational costs	2600	1	1297	1
<i>Total input cost</i>	8301		4818 (58%)	

As shown in Table 7.11 in order to achieve economic efficiency with minimal changes to the TE input levels, the number of branch/account managers must be reduced to approximately 69% of current levels. The number of administrative/commercial staff and tellers only need to be reduced to 80% and 86% of current levels, respectively. The 'implicit prices' are on average 20% below the current levels. The next section compares Farrell CE targets (assuming fixed prices) with economic efficient targets (assuming variable prices).

7.4.3 Comparison of CE measures

7.4.3.1 Overall network level

The summary statistics of Farrell CE targets and economic efficient targets are shown in Table 7.12. The input targets are reported as the total for the network. The prices targets are reported as the average for all branches.

Table 7.12 - Comparison of targets at the overall network level

	<i>Initial values</i>		<i>Farrell CE</i>		<i>Economic efficiency</i>			
	<i>Inputs</i>	<i>Prices</i>	<i>Inputs</i>	<i>Fixed prices</i>	<i>Inputs</i>	<i>Minimal prices</i>	<i>Inputs</i>	<i>Implicit prices</i>
No. managers	505	5.38	341	5.38	336	4.36	346	4.12
No. adm/com staff	477	3.37	386	3.37	404	2.22	384	2.83
No. tellers	335	4.22	291	4.22	286	3.23	288	3.63
Operational costs	2600	1	1565	1	1561	1	1297	1
<i>Total cost</i>	<i>8301</i>		<i>5912</i>		<i>4818</i>			

From Table 7.12 it can be concluded that with current prices considered fixed, the total cost can only be reduced to 71% of the current level, under the existing technological constraints. With variable prices the total cost can be reduced considerably further to 58% of current levels.

The ‘implicit prices’ are, on average, 20% below current prices, but at the network level the target input levels are quite similar under both assumptions of fixed and variable prices.

7.4.3.2 Individual branch level

This section discusses the implications of the price assumptions on the targets set at the individual branch level. The case of a branch located in the Lisbon region is analysed in detail. Table 7.13 shows the target input and price levels considering scenarios of fixed prices (Farrell CE targets) and variable prices (economic efficient targets).

Table 7.13 - Comparison of targets at the branch level

	<i>Initial values</i>		<i>Farrell CE</i>		<i>Economic efficiency</i>			
	<i>Inputs</i>	<i>Prices</i>	<i>Inputs</i>	<i>Fixed prices</i>	<i>Inputs</i>	<i>Minimal prices</i>	<i>Inputs</i>	<i>Implicit prices</i>
No. managers	5	5.53	4.4	5.53	3.7	4.56	4.5	3.55
No. adm/com staff	4	3.34	4.4	3.34	5.9	1.54	3.9	2.97
No. tellers	3	4.39	3.2	4.39	2.7	3.78	3.4	4.02
Operational costs	22.0	1	16.9	1	18.5	1	13.4	1
<i>Total cost</i>	<i>76.1</i>		<i>69.9</i>		<i>54.7</i>			

Table 7.13 shows that considering the prices fixed, the costs can only be reduced to 92% (i.e., 69.9/76.1) of current levels. However, if allowing for price adjustments, the efficient cost is only 72% (i.e., 54.7/76.1) of the current level. This indicates that a substantial amount of the cost incurred by this branch is due to having current prices above the minimal price levels observed in its region (i.e., price inefficiency).

In order to achieve economic efficiency the branch can either adopt the minimal regional prices¹⁹ or the ‘implicit prices’. The main advantage of the targets based on the ‘implicit

¹⁹ Note that the minimal regional prices reported in Table 7.13 leading to economic efficiency are similar to those reported in Table 7.5 for the Lisbon region.

prices' is that the resulting target input mix is closer to the original balance between the inputs, which facilitates the implementation of the changes required to attain efficiency.

The negative side of the target setting method based on 'implicit prices' is that some of the prices suggested may be below the observed minimal prices. In the case of this branch, the results suggest that managers should be paid, on average, 22% below the minimal regional salary. This is necessary in order to have a target level of managers of 4.5 (closer to the current 5) instead of 3.7 (as required if the managers are paid at the minimal market prices). If paying salaries below the minimal levels observed in other branches is considered unattainable, then the number of managers has to be reduced further.

The targets based on 'implicit prices' have the advantage of enabling paying a salary that reflects adequately the relative productivity of employees with different functions. Note that for this branch the administrative/commercial staff and tellers could be paid above the minimal regional salary.

It should also be noted that ideally all inputs should be measured in physical quantities. However, due to data unavailability, the inputs are often measured by their monetary value, as was the case for other operational costs (excluding staff costs) in this assessment. When this happens, the target input should be multiplied by its corresponding 'implicit price' in order to obtain a meaningful cost target (i.e. the value for other operational costs of 13.4 reported in Table 7.13 results from multiplying the target input of 17.2 by a target price of 0.78).

7.5 Summary and conclusions

This chapter explored the assessment of cost efficiency in complex scenarios of price uncertainty and in situations where price adjustments are possible.

Its main contribution can be stated both in terms of the novel developments to the DEA method and the managerial implications of enhancing CE assessments to include broader assumptions regarding input prices. The models and methods were developed considering an assessment under constant returns to scale.

The first scenario considered situations of price uncertainty at the DMU level. This chapter described a method for the estimation of a confidence interval for the cost efficiency measure. This involved the development of an alternative formulation of the DEA model, which measures CE under an ‘optimistic’ and ‘pessimistic’ perspective in relation to the potential input prices faced by each DMU.

The results of the case study of the bank branch network showed that the CE measures obtained with ‘optimistic’ and ‘pessimistic’ perspectives do not differ significantly from the CE measure obtained using the exact prices paid at each DMU. This indicates that the DEA method can provide robust estimates of CE even with price uncertainty.

The second scenario analysed considered cases where the DMs have some discretion in defining the input prices at the DMUs. In such circumstances, a new efficiency component should be considered in the evaluation of economic efficiency. This component, called ‘price efficiency’, captures the extent to which costs are increased due to paying prices above the minimal levels observed in other DMUs operating in the same market.

In order to become fully efficient, in the sense of producing current output levels at the minimal cost, the DMUs may have to adjust both input and price levels. In order to facilitate the implementation of these minimal cost targets, this chapter developed a target setting method which enables keeping the input balance of each DMU as close as possible to the original input mix.

This involves adopting price targets that reflect the marginal rate of substitution (i.e. relative productivity) between the inputs. In order to ensure the feasibility of these price targets, their relative magnitude must be within the relative bounds observed in other DMUs' facing similar market conditions. These targets enable an elimination of inefficiencies as smooth as possible for the organisation, avoiding reallocations of resources and significant changes to the operating practices. The major limitation of this method is that some inputs may be set price targets below the minimal levels observed in the market, in exchange for keeping the associated input level close to the TE value. This may involve negotiation of employment conditions rather than discharging staff.

In terms of the results of the analysis of the bank branch network allowing for adjustment of input prices, the results indicated that the major cause of cost increase is due to paying salaries above the minimal values observed in other branches operating in the same market. The network costs could be reduced to 58% of current levels through the elimination of both input and price related inefficiencies. The targets based on 'implicit prices' suggested a reduction of branch/account managers to 69% of current levels, whereas the number of administrative/commercial staff and tellers need only a reduction to 80% and 86% of current levels, respectively. In addition, the attainment of price

efficiency would require reductions to salaries that could be around 20% in some branches.

These results indicated the presence of considerable cost inefficiencies in this network. An explanation for the presence of such large inefficiencies may be the fact that cost efficiency is not the sole driver in the management of the operational activity of branch networks. Other targets, related to marketing issues, quality of service, and financial gains from the intermediation activity may be given priority over the achievement of cost efficiency in branches' operational activity.

CHAPTER 8

Assessment of bank branches' performance at the regional level: Comparison of regional efficiency spread and frontier productivity

8.1 Introduction

This chapter develops a method for the comparison of performance of groups of DMUs operating under different conditions, at the same time period. The performance measure obtained for each group can be separated into two components. One is attributable to the policies or conditions within which the DMUs are required to operate. The other component reflects managerial inefficiency.

The new method described in this chapter is based on the Malmquist index, introduced by Caves *et al.* (1982) and developed further by Fare, Grosskopf, Lindgren and Roos (1994). The Malmquist index is usually applied to the measurement of productivity change over time and can be decomposed into an efficiency change index and a technological change index. Similarly, the performance index developed in this chapter can be decomposed into an index reflecting the efficiency spread among DMUs operating in similar conditions, and an index of the productivity gap between the “best-practice” frontiers of the different groups. The developments described in this chapter are an extension to the Malmquist-type indexes used in Berg *et al.* (1993) and Pastor *et al.* (1997) in the context of the evaluation of performance of financial institutions.

The method developed in this chapter is an alternative to the 'program efficiency' approach developed by Charnes *et al.* (1981), which enables the comparison between frontiers corresponding to different programs or conditions underlying the DMUs' activity.

In relation to the performance assessment of the bank branch network used as the case study for this thesis, the analysis reported in the previous chapters did not consider the effect of the environmental conditions on branches' performance due to data unavailability at the DMU level. Therefore, the aim of this section is to complement the previous analysis by characterising the performance profile of branches in different regions. This is a way to identify the relative strengths and weaknesses of each region and to distinguish inefficiencies that can be attributable to poor management from those that come with the market (i.e., caused by less favourable environmental conditions).

This chapter unfolds as follows. Section 8.2 reviews the 'program efficiency' method (Charnes *et al.*, 1981), which enables comparing the performance of DMUs operating under different programs or conditions. The limitations of this method are discussed, motivating the development of the approach described in this chapter. Section 8.3 introduces the Malmquist index for productivity measurement and its recent developments. Section 8.4 develops a new performance index for the comparison of groups of DMUs operating under different conditions. Section 8.5 illustrates the computation of the distance functions underlying the construction of the index using the DEA approach. Section 8.6 describes the data used, and section 8.7 presents the empirical results. Section 8.8 summarises and concludes.

8.2 Review of the 'program efficiency' method

Charnes *et al.* (1981) were the first to propose a method to separate the DEA efficiency measure into two components, one only attributable to the context within which the DMUs are required to operate (i.e., associated with programs, policies or environmental conditions) and the other reflecting managerial inefficiency.

In order to describe the method proposed by Charnes *et al.* (1981), known as 'program efficiency', consider a case with k ($p = 1, \dots, k$) different programs, each with n_p DMUs. For each set of n_p DMUs, a DEA assessment can be run in the usual manner by solving any of the models described in Chapter 2. This assessment yields a measure of the managerial efficiency of each DMU within program p .

The programs can then be compared by eliminating the managerial inefficiencies, projecting each DMU to the frontier associated with its own program. A second DEA assessment is then carried out comprising the DMUs from all programmes, with input-output levels free of managerial inefficiencies. The inefficiencies observed at this stage, measured against a pooled frontier, are only attributable to the program characteristics, rather than to managerial sources, as all the DMUs included in the assessment are efficient within their specific program.

The ideas involved are illustrated in Figure 8.1. The DMUs used in this example produce equal amounts of one output using two different inputs (x_1 and x_2). These DMUs operate under two different programs, represented in Figure 8.1 using different symbols.

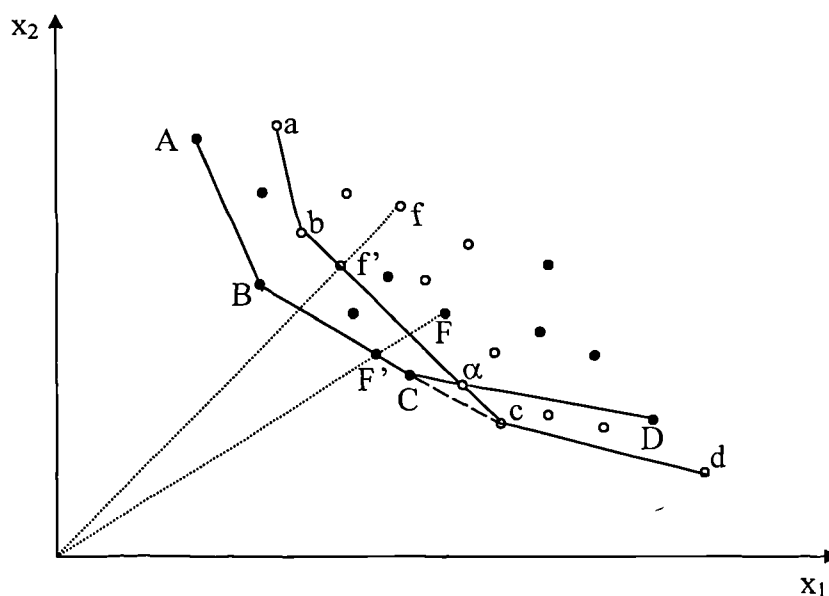


Figure 8.1 – Illustration of the 'Program Efficiency' method

The first stage of the 'program efficiency' method involves the measurement of relative efficiency of DMUs operating in the same program. Therefore, efficiency is measured against a group-specific frontier. It reflects exclusively the managerial efficiency of the DMUs. In Figure 8.1, the piece-wise segments 'abcd' and 'ABCD' define the group-specific frontiers.

The second stage of the method consists of removing managerial (or within-program) inefficiency by adjusting the input-output levels of the DMUs to the frontier of their own group. This is illustrated in Figure 8.1 by the projection of DMUs f and F to their respective frontiers, corresponding to points f' and F' . This is followed by a second DEA assessment, including all DMUs from the different programs. The piece-wise segment $ABCcd$ defines the pooled frontier underlying this assessment. The inefficiencies detected at this stage are only attributable to the different programs within which the DMUs operate rather than to managerial causes, as managerial inefficiencies were eliminated previously by adjusting the data to within-program efficient levels.

Some examples of the use of this method in the assessment of financial institutions can be found in Elyasiani and Mehdiian (1995) and Grifell-Tatjé and Lovell (1997).

The major limitations of the 'program efficiency' method are that the programs are only compared in terms of the productivity of the best-practice frontiers, without it being possible to reflect the efficiency spread found within each program. In addition, some parts of the pooled frontier may result from convex combinations of DMUs from different programs (as the segment between C and c in Figure 8.1)¹. However, there is no empirical evidence that the production possibility set defined in this way is attainable.

The next section develops a method that can compare the performance of programs including both effects of efficiency spread and frontier productivity. In addition it avoids the areas of misspecification introduced by segments of the frontier such as the one defined by points C and c in Figure 8.1.

8.3 The Malmquist index

In recent years the Malmquist index has become the standard approach to productivity measurement within the non-parametric literature. Malmquist indexes were introduced by Caves *et al.* (1982). They named these indexes after Malmquist, who had earlier proposed constructing input quantity indexes as ratios of distance functions. The concept of a distance function is explained next.

¹ Note that the DEA method cannot construct a pooled frontier defined by $ABC\alpha cd$, as this frontier would be non-convex.

In 1953, independently of each other, Shephard (1953) and Malmquist (1953) introduced the distance function as a tool in economics. Shephard mainly used it for duality theory, while Malmquist applied it to index number theory. To introduce the concept of a distance function, consider that in time period t the DMUs are using inputs $X^t \in \mathbb{R}_+^m$ to produce outputs $Y^t \in \mathbb{R}_+^s$. The technology of production Φ^t can be defined as follows:

$$\Phi^t = \left\{ (X^t, Y^t) \mid \text{Input vector } X^t \text{ can produce the output vector } Y^t \right\}. \quad (8.1)$$

It consists of all input-output vectors that are technically feasible for a certain production process.

The input distance function is defined on the technology Φ^t as the maximal feasible contraction of X^t that still enables producing Y^t , as follows:

$$D_i(X^t, Y^t) = \sup \left\{ \lambda : \left(\frac{X^t}{\lambda}, Y^t \right) \in \Phi^t \right\}. \quad (8.2)$$

The input distance function is the reciprocal to Farrell's (1957) input-oriented measure of technical efficiency (see Fare and Lovell, 1978). Note that $D_i(X^t, Y^t) \geq 1$ if and only if $(X^t, Y^t) \in \Phi^t$.

To define a Malmquist index requires specification of two mixed-period distance functions, such as:

$$D_i^t(X^{t+1}, Y^{t+1}) = \sup \left\{ \lambda : \left(\frac{X^{t+1}}{\lambda}, Y^{t+1} \right) \in \Phi^t \right\}; \quad (8.3)$$

$$D_i^{t+1}(X^t, Y^t) = \sup \left\{ \lambda : \left(\frac{X^t}{\lambda}, Y^t \right) \in \Phi^{t+1} \right\}. \quad (8.4)$$

The first mixed-period distance function measures the maximal proportional reduction to inputs required to make a DMU in time period $t+1$, i.e. with (X^{t+1}, Y^{t+1}) , efficient in relation to technology at the previous period t , i.e. in Φ^t .

Similarly, the second mixed-period distance function measures the maximal proportional reduction to inputs required to make a DMU in time period t , i.e. (X^t, Y^t) , efficient in relation to technology at $t+1$, i.e. in Φ^{t+1} . However, since an input-output combination observed in one period may not be feasible within the technology in another period, in both these mixed-period assessments the value of the input distance function may be smaller than unity.

Caves *et al.* (1982) define an input-based Malmquist productivity index relative to a single technology Φ^t (in (8.5)) and Φ^{t+1} (in (8.6)), as follows:

$$M_i^t = \frac{D_i^t(X^{t+1}, Y^{t+1})}{D_i^t(X^t, Y^t)}; \quad (8.5)$$

$$M_i^{t+1} = \frac{D_i^{t+1}(X^{t+1}, Y^{t+1})}{D_i^{t+1}(X^t, Y^t)}. \quad (8.6)$$

The values of M_i^t and M_i^{t+1} may be smaller, equal or greater than one, depending on whether productivity growth, stagnation or productivity decline has occurred between periods t and $t+1$. In general, M_i^t and M_i^{t+1} yield different productivity numbers since their reference technologies may differ.

The Malmquist index was treated as a theoretical one until its enhancement by Fare, Grosskopf, Lindgren and Roos (1994). A major contribution of this paper was to relax the efficiency assumption and provide DEA models for the calculation of the Malmquist index.

Fare, Grosskopf, Lindgren and Roos (1994) defined an input-oriented productivity index as the geometric mean of the two Malmquist indexes referring to the technology at time periods t and $t+1$ (i.e., (8.5) and (8.6)), yielding the following Malmquist-type measure of productivity:

$$M_i^{t,t+1} = \left[\frac{D_i^t(X^{t+1}, Y^{t+1})}{D_i^t(X^t, Y^t)} \cdot \frac{D_i^{t+1}(X^{t+1}, Y^{t+1})}{D_i^{t+1}(X^t, Y^t)} \right]^{1/2}. \quad (8.7)$$

Another major achievement of Fare, Grosskopf, Lindgren and Roos (1994) was to show how to decompose the index $M_i^{t,t+1}$ into an index of technical efficiency change and an index reflecting the change in the frontier of the production possibility set, i.e. an index of technical (or technological) change. These components are obtained by rewriting the index in (8.7) as follows:

$$M_i^{t,t+1} = \frac{D_i^{t+1}(X^{t+1}, Y^{t+1})}{D_i^t(X^t, Y^t)} \cdot \left[\frac{D_i^t(X^{t+1}, Y^{t+1})}{D_i^{t+1}(X^{t+1}, Y^{t+1})} \cdot \frac{D_i^t(X^t, Y^t)}{D_i^{t+1}(X^t, Y^t)} \right]^{1/2}. \quad (8.8)$$

The ratio outside the bracket measures the **input technical efficiency change** between time periods t and $t+1$. The geometric mean of the two ratios inside the bracket captures the **technological change** (or shift in technology) between the two periods, evaluated at the input-output levels at t , i.e. (X^t, Y^t) , and at $t+1$, i.e. (X^{t+1}, Y^{t+1}) .

In order to illustrate pictorially the concepts underlying this index, consider the case of production with two inputs (x_1 and x_2) and only one output at time periods t and $t+1$, as shown in Figure 8.2. The technology is drawn such that $\Phi^{t+1} \supseteq \Phi^t$.

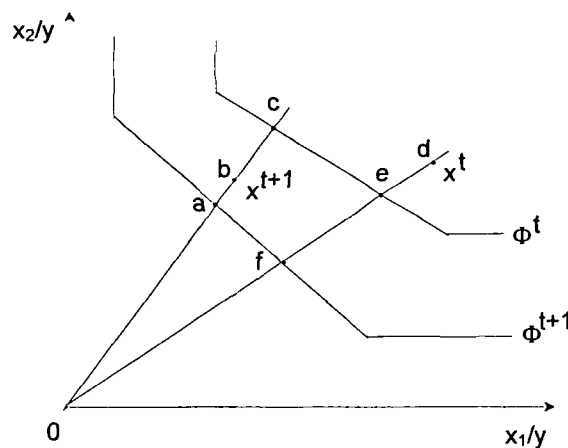


Figure 8.2 - Malmquist input-based productivity index

In terms of distances, the productivity index $M_i^{t,t+1}$ can be written as follows:

$$M_i^{t,t+1} = \frac{ob/oa}{od/oe} \cdot \left[\frac{ob/oc}{ob/oa} \cdot \frac{od/oe}{od/of} \right]^{1/2} \quad (8.9)$$

Rearranging in terms of efficiency measures:

$$M_i^{t,t+1} = \frac{oe/od}{oa/ob} \cdot \left[\frac{oa}{oc} \cdot \frac{of}{oe} \right]^{1/2} \quad (8.10)$$

The term outside the bracket in (8.10) represents the ratio of technical efficiency at time periods t and $t+1$, capturing changes in efficiency over time. A value less than unity indicates an efficiency improvement over time.

The ratios inside the bracket measure shifts in technology at input levels x^t and x^{t+1} , and so technological change is measured as the geometric mean of the frontier shifts for the input levels at x^t and x^{t+1} . Again a value less than unity indicates an improvement in technological conditions.

Overall, improvements in productivity yield input Malmquist indexes with values smaller than unity. Note that productivity growth may involve technological regress (if

gains in efficiency dominate the regress in the frontier of the PPS from t to $t+1$) or a fall in efficiency (if technological progress in the frontier of the PPS from t to $t+1$ dominates the loss in DMUs' efficiency). Similar possibilities hold for the case of decline in productivity.

Recent studies using the Malmquist index revealed many instances of intersecting annual technologies (such that $\Phi^t \not\subset \Phi^{t+1}$), which suggested the presence of non-neutral technological change. Non-neutral (or biased) technological change occurs when the shift in the frontier depends on the operating mix of the DMUs.

However, the original decomposition of the Malmquist index contained no index reflecting the bias of technological change. This has led Fare *et al.* (1997) [see also Grifell-Tatjé and Lovell (1997)] to propose the decomposition of the technological change component of the Malmquist index into an index of magnitude of technological change and an index of bias of technological change. This decomposition is based on the use of the original Caves *et al.* (1982) index relative to a single period technology Φ^t , as in (8.5), as follows:

$$\begin{aligned}
 M_i^t &= \frac{D_i^t(X^{t+1}, Y^{t+1})}{D_i^t(X^t, Y^t)} \\
 \Leftrightarrow M_i^t &= \frac{D_i^{t+1}(X^{t+1}, Y^{t+1})}{D_i^t(X^t, Y^t)} \cdot \frac{D_i^t(X^{t+1}, Y^{t+1})}{D_i^{t+1}(X^{t+1}, Y^{t+1})} \\
 \Leftrightarrow M_i^t &= \frac{D_i^{t+1}(X^{t+1}, Y^{t+1})}{D_i^t(X^t, Y^t)} \cdot \frac{D_i^t(X^t, Y^t)}{D_i^{t+1}(X^t, Y^t)} \cdot \left[\frac{D_i^t(X^{t+1}, Y^{t+1})}{D_i^{t+1}(X^{t+1}, Y^{t+1})} \cdot \frac{D_i^{t+1}(X^t, Y^t)}{D_i^t(X^t, Y^t)} \right] \quad (8.11)
 \end{aligned}$$

The first ratio in (8.11) measures the technical efficiency change. The second ratio measures the **magnitude of technological change** along a ray through period t data. This term provides a local measure of the magnitude of technological progress or

technological regress. The component inside the bracket measures the **bias of technological change** as the ratio of the magnitude of technological change along a ray through period $t+1$ data to the magnitude of technological change along a ray through period t data. If technological change is neutral, the production possibility set shifts out (or in) by the same proportion along a ray through period $t+1$ data as it does along a ray through period t data. If technological change is biased, the production possibility set shifts out (or in) by a different proportion along a ray through period $t+1$ data than it does along a ray through period t data. This bias component makes a positive or negative contribution to productivity change as management adjusts the input mix of production in the right or wrong direction when confronted with a non-neutral change in the production possibility set. Its value is smaller than unity when management re-orientes production in the right direction in response to changing technology, and greater than or equal to unity otherwise.

Another recent addition to the theory underlying the construction of the Malmquist index relates to the returns to scale assumption that should be used for its calculation. Grifell-Tatjé and Lovell (1995) reported an example which showed that the Malmquist index provides an inaccurate productivity measure when it is evaluated under variable returns to scale (VRS). This was further discussed by Fare and Grosskopf (1996), who proposed the use of constant returns to scale (CRS) as the reference technology for calculating the overall Malmquist index, regardless of the form of the real technology underlying the DMUs' activity. The deviations from CRS may be identified through a further decomposition of the efficiency change component into **scale efficiency change** and **pure technical efficiency change** components. The latter is measured relative to the variable returns to scale technology. In this case we have:

$$\text{Efficiency change: } \frac{S_i^t(X^t, Y^t)}{S_i^{t+1}(X^{t+1}, Y^{t+1})} \cdot \frac{D_i^{t+1}(X^{t+1}, Y^{t+1}|V)}{D_i^t(X^t, Y^t|V)}. \quad (8.12)$$

In the formula above, the notation $D_i^t(X^t, Y^t|V)$ corresponds to an input distance function defined on a VRS technology at time period t , evaluated at the input-output levels at t , i.e. (X^t, Y^t) . Recall that $D_i^t(X^t, Y^t)$ refers to an evaluation considering a CRS technology at time period t . $S_i^t(X^t, Y^t)$ corresponds to an input oriented measure of scale efficiency, defined as:

$$S_i^t(X^t, Y^t) = D_i^t(X^t, Y^t|V) / D_i^t(X^t, Y^t). \quad (8.13)$$

The previous description of the Malmquist index only considered two time periods under analysis. In a more general case where more than two periods of cross-sectional data are available, one desirable property of the productivity evaluation is that it should satisfy the circular test, as follows (see Fare and Grosskopf, 1996):

$$\blacksquare \quad M_i^{t,t+1} \times M_i^{t+1,t+2} = M_i^{t,t+2}. \quad (8.14)$$

If the index satisfies the circular test, it is possible to decompose the productivity changes over the whole period into sub-periods in a consistent way.

For a comprehensive review of the literature on the theoretical developments and applications on the Malmquist index see Fare *et al.* (1998).

8.4 New method for comparing the performance of DMUs operating under different conditions

This section develops an index for the comparison of performance of DMUs facing different operating conditions. It is illustrated within the context of the comparison of performance of bank branches operating in different regions.

The index proposed is an adaptation to the Malmquist index such that it no longer measures the productivity change between two time periods. The new index corresponds to a cross-sectional comparison of performance of groups of DMUs operating in different conditions, at a certain moment in time. This new performance index can be multiplicatively decomposed into an index reflecting the efficiency spread within each group and an index reflecting the differences in productivity between the best-practice frontiers.

To illustrate the derivation of this new index, consider α DMUs in region A, using inputs $X^A \in \mathfrak{R}_+^m$ to produce outputs $Y^A \in \mathfrak{R}_+^s$, and β DMUs in region B, using inputs $X^B \in \mathfrak{R}_+^m$ to produce outputs $Y^B \in \mathfrak{R}_+^s$. The DMUs operating in region A are represented by their input-output vector as $(X_{j_a}^A, Y_{j_a}^A)$ for $j_a = 1, \dots, \alpha$. A similar notation is used for region B.

We shall first discuss the component of the index that compares **within-region efficiency spread** (IE_i^{AB}). Its value is given by the ratio of the geometric means of the distance of the DMUs to their regional specific frontier, as follows:

$$IE_i^{AB} = \frac{\left[\prod_{j_a=1}^{\alpha} D_i^A(y_{j_a}^A, x_{j_a}^A) \right]^{1/\alpha}}{\left[\prod_{j_b=1}^{\beta} D_i^B(y_{j_b}^B, x_{j_b}^B) \right]^{1/\beta}}. \quad (8.15)$$

A value of IE_i^{AB} less than one indicates that the efficiency spread is smaller (i.e., there is greater homogeneity) in DMUs of region A than in region B.

Note that the aggregation of the efficiency scores of the DMUs with respect to their specific PPS frontier is in line with the notion of an industry efficiency score, described

in Farrell (1957). In the context of the regional comparison discussed in this section, the industry may be regarded as a group of DMUs operating under similar environmental conditions. In this case, the index IE_i^{AB} above can be interpreted as a comparison of industry efficiency between two regions.

In relation to the returns to scale assumption used for the estimation of the distance functions, for the motives exposed in the review on the previous section, constant returns to scale should be used. In case the DMUs operate under variable returns to scale, the within-region efficiency spread index can be decomposed into a scale efficiency component and a pure technical efficiency component. The resulting measures are as follows:

$$IE_i^{AB} = \frac{\left[\prod_{j_b=1}^{\beta} S_i^B(X_{j_b}^B, Y_{j_b}^B) \right]^{1/\beta} \left[\prod_{j_a=1}^{\alpha} D_i^A(X_{j_a}^A, Y_{j_a}^A | V) \right]^{1/\alpha}}{\left[\prod_{j_a=1}^{\alpha} S_i^A(X_{j_a}^A, Y_{j_a}^A) \right]^{1/\alpha} \left[\prod_{j_b=1}^{\beta} D_i^B(X_{j_b}^B, Y_{j_b}^B | V) \right]^{1/\beta}}. \quad (8.16)$$

The first term in (8.16) enables a comparison of within-region scale efficiency between regions A and B. The second term compares the pure technical efficiency between the two regions.

The other component of the index compares the **regional frontier productivity** (i.e., the difference in productivity between the regional best-practice frontiers). Recall that when measuring the shift in technology between two time periods, the component of the Malmquist index measuring technological change is DMU-specific and it is evaluated at the input-output levels of the same DMU at time periods t and $t+1$ (see formula (8.8)). When comparing regional frontiers at the same moment in time, their distance is measured at the input-output levels of all DMUs in both regions, and the values are

aggregated using the geometric mean. The resulting component of the index for measuring the distance in best-practice frontiers of two regions A and B is given by:

$$IF_i^{AB} = \left[\prod_{j_a=1}^{\alpha} \frac{D_i^B(X_{j_a}^A, Y_{j_a}^A)}{D_i^A(X_{j_a}^A, Y_{j_a}^A)} \cdot \prod_{j_b=1}^{\beta} \frac{D_i^B(X_{j_b}^B, Y_{j_b}^B)}{D_i^A(X_{j_b}^B, Y_{j_b}^B)} \right]^{1/(\alpha+\beta)} \quad (8.17)$$

A value of IF_i^{AB} less than one indicates greater productivity (or dominance) of the frontier of region A with respect to region B. $D_i^A(X_{j_b}^B, Y_{j_b}^B)$ represents the input distance function for a DMU in region B with respect to the frontier of region A. Note that both within-region and inter-region distance functions are needed for the calculation of the index.

The product of the two indexes in (8.15) and (8.17) gives the **overall index** for the comparison of performance of groups of DMUs operating in different regions, as follows:

$$I_i^{AB} = \frac{\left[\prod_{j_a=1}^{\alpha} D_i^A(X_{j_a}^A, Y_{j_a}^A) \right]^{1/\alpha}}{\left[\prod_{j_b=1}^{\beta} D_i^B(X_{j_b}^B, Y_{j_b}^B) \right]^{1/\beta}} \cdot \left[\prod_{j_a=1}^{\alpha} \frac{D_i^B(X_{j_a}^A, Y_{j_a}^A)}{D_i^A(X_{j_a}^A, Y_{j_a}^A)} \cdot \prod_{j_b=1}^{\beta} \frac{D_i^B(X_{j_b}^B, Y_{j_b}^B)}{D_i^A(X_{j_b}^B, Y_{j_b}^B)} \right]^{1/(\alpha+\beta)} \quad (8.18)$$

A value less than unity indicates better performance in region A than in region B. This may be due to two causes: less dispersion in efficiency levels of DMUs in region A than in region B, or the dominance of the best-practice frontier of region A with respect to the frontier of region B.

When it is possible to specify a “typical” DMU for each region, represented by (X_{typ}^A, Y_{typ}^A) and (X_{typ}^B, Y_{typ}^B) , then the index in (8.18) reduces to a formulation similar to the usual Malmquist index in (8.8), as follows²:

$$I_i^{AB(typ)} = \frac{D_i^A(X_{typ}^A, Y_{typ}^A)}{D_i^B(X_{typ}^B, Y_{typ}^B)} \left[\frac{D_i^B(X_{typ}^A, Y_{typ}^A)}{D_i^A(X_{typ}^A, Y_{typ}^A)} \frac{D_i^B(X_{typ}^B, Y_{typ}^B)}{D_i^A(X_{typ}^B, Y_{typ}^B)} \right]^{1/2}. \quad (8.19)$$

Based on this type of approach, Berg *et al.* (1993) reported comparisons of productivity across banks in the Nordic countries. The comparisons were made between the largest banks of each country and between the average banks. Pastor *et al.* (1997) compared the relative efficiency and differences in technology of several European and the US banking systems using a Malmquist-type index as above. They reported the results for the alternative definitions of a “typical” bank, based on the median bank, the simple average of banks and the weighted (by assets) average of banks. The advantage of the new index developed in this section is that it does not require a subjective definition of a “typical” DMU, as it can handle directly all the observations corresponding to individual DMUs.

The previous description of the performance index only considered a comparison between two regions. In a more general setting, where it is important to obtain a comparison between more than two regions, it is desirable that the performance comparison satisfies the circular test (see Fare *et al.*, 1996), as follows:

$$\blacksquare \quad I_i^{AB} \times I_i^{BC} = I_i^{AC}. \quad (8.20)$$

² Note that in the index in (8.19) the frontier of region A can be seen as the frontier at time $t+1$, and the frontier of region B corresponds to the frontier at time t . Similarly, (X_{typ}^A, Y_{typ}^A) corresponds to a DMU at time period $t+1$ and (X_{typ}^B, Y_{typ}^B) corresponds to a DMU at time period t .

The equality above implies that the index for the comparison of performance between two regions (I_i^{AC}) can be obtained by comparing these regions (A and C) to any other region considered as the reference underlying the comparison (region B in formula (8.20)). This property ensures that the ranking of regional performance is independent of the reference used for the calculation of the index.

If the circular test is verified, it ensures that the index is transitive, such that:

$$\blacksquare \text{ if } I_i^{AB} < 1 \text{ and } I_i^{BC} < 1, \text{ then } I_i^{AC} < 1 \quad (8.21)$$

The proposition (8.21) indicates that if region B has a larger efficiency spread than region A, and region C has a larger efficiency spread than region B, then region C has a larger efficiency spread than region A.

This proposition can be easily proved. Starting from the statement that $I_i^{AB} < 1$, then multiplying the left-hand side of the inequality by I_i^{BC} , if the value of I_i^{BC} is also smaller than unity, we obtain $I_i^{AB} \times I_i^{BC} < 1$. From the equality in (8.20) we conclude that $I_i^{AC} < 1$.

In relation to the efficiency spread index in (8.15), the circular test is satisfied, as follows:

$$\blacksquare \quad IE_i^{AB} \times IE_i^{BC} = IE_i^{AC} . \quad (8.22)$$

This equality in (8.22) can be easily verified:

$$\frac{\left[\prod_{j_a=1}^{\alpha} D_i^A(y_{j_a}^A, x_{j_a}^A) \right]^{1/\alpha}}{\left[\prod_{j_b=1}^{\beta} D_i^B(y_{j_b}^B, x_{j_b}^B) \right]^{1/\beta}} \times \frac{\left[\prod_{j_b=1}^{\beta} D_i^B(y_{j_b}^B, x_{j_b}^B) \right]^{1/\beta}}{\left[\prod_{j_c=1}^{\chi} D_i^C(y_{j_c}^C, x_{j_c}^C) \right]^{1/\chi}} = \frac{\left[\prod_{j_a=1}^{\alpha} D_i^A(y_{j_a}^A, x_{j_a}^A) \right]^{1/\alpha}}{\left[\prod_{j_b=1}^{\chi} D_i^C(y_{j_b}^C, x_{j_b}^C) \right]^{1/\chi}} .$$

This indicates that the index comparing the efficiency spread of two regions (IE_i^{AC}) can be obtained by comparing these regions to any other region considered as the reference for the comparison (e.g., B).

In relation to the frontier productivity index in (8.17), the circular test is not satisfied, i.e:

$$\blacksquare IF_i^{AB} \times IF_i^{BC} \neq IF_i^{AC}. \quad (8.23)$$

The index comparing the productivity of the frontiers of regions A and C (IF_i^{AC}) cannot be obtained from the indexes based on comparisons to a reference region B (IF_i^{AB} and IF_i^{BC}) because this would involve the evaluation of frontier distances at the input-output levels observed in DMUs from the three regions involved (A, B and C). Conversely, the index IF_i^{AC} only evaluates the distance between the frontiers at the input-output levels of DMUs in regions A and C.

As a result, the index reflecting the distance between the regional frontiers depends on the region underlying the comparison. Thus, the frontier productivity ranking is also affected by the choice of the reference region.

In order to obtain an index that measures the distance between the regional frontiers and also satisfies the circular test, the distance between any two frontiers must be evaluated at the input-output levels of all DMUs in the regions under comparison.

For example, in case the ranking involves four different regions (e.g., A, B, C and D, with α , β , χ and δ DMUs in each region), the circular index for the comparison of frontier productivity between regions A and B would be as follows:

$$IF(adj)_i^{AB} = \left[\prod_{j_a=1}^{\alpha} \frac{D_i^B(y_{j_a}^A, x_{j_a}^A)}{D_i^A(y_{j_a}^A, x_{j_a}^A)} \cdot \prod_{j_b=1}^{\beta} \frac{D_i^B(y_{j_b}^B, x_{j_b}^B)}{D_i^A(y_{j_b}^B, x_{j_b}^B)} \cdot \prod_{j_c=1}^{\gamma} \frac{D_i^B(y_{j_c}^C, x_{j_c}^C)}{D_i^A(y_{j_c}^C, x_{j_c}^C)} \cdot \prod_{j_d=1}^{\delta} \frac{D_i^B(y_{j_d}^D, x_{j_d}^D)}{D_i^A(y_{j_d}^D, x_{j_d}^D)} \right]^{1/(\alpha+\beta+\gamma+\delta)} \quad (8.24)$$

For the index in (8.24), the circular test is verified:

$$\blacksquare \quad IF(adj)_i^{AB} \times IF(adj)_i^{BC} = IF(adj)_i^{AC}. \quad (8.25)$$

The use of this adjusted index ensures that the comparison of frontier distances is independent of the reference region chosen.

Finally, the overall performance index that should be used for the comparison of performance of more than two regions is obtained as follows³:

$$I(adj)_i^{AB} = IE_i^{AB} \times IF(adj)_i^{AB}. \quad (8.26)$$

This index also satisfies the circular test and enables obtaining a robust performance ranking of groups of DMUs operating under different programs or conditions.

8.5 Computation of distance functions using the DEA approach

The work by Fare and Lovell (1978) showed that the distance function is the reciprocal to Farrell's (1957) measure of technical efficiency. This opened the possibility of using DEA models to compute the Malmquist index.

The analysis of productivity change over time using the Malmquist index requires the calculation of within-period and mixed-period distance functions. In the context of comparison of regional performance, the performance index developed in this chapter

³ Note that as IF_i^{AB} is not circular, the overall performance index I_i^{AB} (equal to $IE_i^{AB} \times IF_i^{AB}$) is not circular either. Therefore, it should not be used for comparisons involving more than two groups of DMUs.

involves evaluating the distance of the DMUs to their own regional frontier, as well as to the frontier of other regions, corresponding to 'within-group' and 'mixed-group' distance functions.

Considering a DMU j_0 located in region A, model (8.27) derives the efficiency with respect to its own regional frontier, whereas model (8.28) derives the efficiency with respect to the frontier of a different region (region B).

Within-region efficiency model		(8.27)
$E_i^A(X_{j_0}^A, Y_{j_0}^A) = \text{Min } \theta$		
s.t. $\sum_{j=1}^{\alpha} x_{ij}^A \lambda_j \leq \theta x_{ij_0}^A$	$i = 1, \dots, m$	
$\sum_{j=1}^{\alpha} y_{rj}^A \lambda_j \geq y_{rj_0}^A$	$r = 1, \dots, s$	
$\lambda_j \geq 0$	$j = 1, \dots, \alpha$	

Inter-region efficiency model		(8.28)
$E_i^B(X_{j_0}^A, Y_{j_0}^A) = \text{Min } \theta$		
s.t. $\sum_{j=1}^{\beta} x_{ij}^B \lambda_j \leq \theta x_{ij_0}^A$	$i = 1, \dots, m$	
$\sum_{j=1}^{\beta} y_{rj}^B \lambda_j \geq y_{rj_0}^A$	$r = 1, \dots, s$	
$\lambda_j \geq 0$	$j = 1, \dots, \beta$	

Note that the efficiency level obtained from model (8.28) involves comparing a DMU of a certain region with the frontier of another region, so that efficiency values greater, equal or smaller than unity can occur. An efficiency level greater than one means that the DMU j_0 in region A is more efficient than any DMU in region B. Therefore, it would be

located outside the boundary of the production possibility set of region B. In the case of efficiency values smaller or equal to one, the DMU would be inside or at the frontier of the production possibility set, respectively.

Both the within-region and inter-region DEA models are solved for every DMU to derive the indexes to compare regional performances. The value of the distance functions are obtained from the efficiency estimates as follows:

$$D_i^A(X_{j_0}^A, Y_{j_0}^A) = 1/E_i^A(X_{j_0}^A, Y_{j_0}^A), \text{ and} \quad (8.29)$$

$$D_i^B(X_{j_0}^A, Y_{j_0}^A) = 1/E_i^B(X_{j_0}^A, Y_{j_0}^A).$$

8.6 Description of the empirical data

The comparison of the regional performances within the network was made focusing on branches' operational activity. The inputs and outputs were defined based on the production approach, as described in Chapters 5 and 7. The variables used in this analysis are listed below:

- Inputs :**
- Number of branch and account managers.
 - Number of administrative and commercial staff.
 - Number of tellers.
 - Operational costs (excluding staff costs).
- Outputs:**
- Total value of deposits.
 - Total value of loans.
 - Total value of off balance sheet business.
 - Number of general service transactions.

The number of branches analysed was 144. These branches are scattered across Portugal, with a greater concentration around the main cities. The network management is divided into four regions, with certain autonomy regarding the definition of branches'

operational procedures and targets. Each region has the following number of branches:

Lisbon: 39, Porto: 24, Centre-South: 47 and North: 34.

The summary statistics for the input-output data in each region are shown in Table 8.1.

Table 8.1 - Summary statistics for the input/output data

Inputs/Outputs	Lisbon		Porto		Centre-South		North	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
No. branch/account managers	4.0	1.5	3.7	1.3	3.0	1.3	3.4	1.4
No. administrative/commercial staff	4.3	2.1	2.5	1.3	3.5	2.2	2.4	1.0
No. tellers	2.6	1.0	1.9	0.7	2.6	1.2	2.0	1.0
Operational costs	22.5	16.8	15.0	4.0	16.5	5.6	17.3	5.3
Value of deposits	4208	3519	3256	3499	2974	2300	4273	2901
Value of loans	1901	1426	1056	606	1050	738	768	579
Value of off-balance sheet business	2035	1993	1125	886	973	676	1234	640
No. of general service transactions	112954	42337	75321	27976	92876	46365	80312	36907

The next section presents the empirical results of the comparison of regional performance using the indexes developed earlier in this chapter.

8.7 Empirical results

This section reports the results of the new index developed for the comparison of regional performance. The components of the index for the comparison of the within-region efficiency spreads and regional frontier productivity are described. These indexes can lead to useful descriptive conclusions about the state of performance within the network.

Table 8.2 presents the component of the index relating to the comparison of within-region efficiency spread (i.e., IE_i^{AB}).

Table 8.2 – Index for the comparison of within-region efficiency spread (IE_i^{AB})

A \ B	Lisbon	Porto	Centre-South	North
Lisbon	1	1.072	0.948	0.998
Porto	0.933	1	0.885	0.931
Centre-South	1.054	1.130	1	1.052
North	1.002	1.074	0.950	1

The results are reported such that a value smaller than unity indicates that the region listed in the row heading has a better efficiency status (i.e., less efficiency spread) than the region listed in the column heading. In terms of formula (8.15) used for the calculation of the values above, the region in the row heading corresponds to region A, whereas the region in the column heading corresponds to B.

Note that the elements below the diagonal of the matrix (shaded in Table 8.2) are the inverse of the associated values in the upper part of the matrix. Also, based on the results reported in Table 8.2, the circularity of the index can be easily verified. For example, the comparison of efficiency spread between Lisbon and the Centre-South can be obtained from the indexes using the North as the reference region:

- $IE_i^{LX,NT} \times IE_i^{NT,CS} = IE_i^{LX,CS}$, i.e., (8.30)
- $0.998 \times 0.950 = 0.948$.

The results of the comparison of the regional efficiency spread are pictorially illustrated in Figure 8.2. Lisbon is considered the reference region, with a value of efficiency spread equal to one (e.g., data from the first row in Table 8.2).

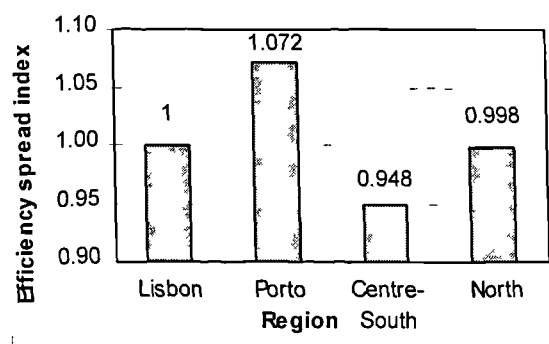


Figure 8.3 – Index for the comparison of within-region efficiency spread, using Lisbon as the reference region

It is concluded that Porto has the smallest efficiency spread⁴, followed by Lisbon, the North and finally the Centre-South, which has the largest efficiency spread.

Table 8.3 reports the results of the component of the index relating to the comparison of productivity between two regional frontiers.

Table 8.3 - *Index for the comparison of productivity between two regional frontiers*
(IF_i^{AB})

A \ B	Lisbon	Porto	Centre-South	North
Lisbon	1	0.817	0.948	0.942
Porto	1.224	1	1.123	1.123
Centre-South	1.062	0.890	1	1.054
North	1.062	0.891	0.949	1

The results in Table 8.3 are reported such that a value smaller than unity indicates that the region listed in the row heading has a frontier with greater productivity than the region listed in the column heading. Note that the elements below the diagonal of the matrix are the inverse of the associated value in the upper part of the matrix.

In the case of this index, the circularity property is not verified, e.g.:

- $IF_i^{LX,NT} \times IF_i^{NT,CS} \neq IF_i^{LX,CS}$, i.e., (8.31)
- $0.942 \times 0.949 \neq 0.948$.

Therefore, this index should not be used for comparisons of more than two regions, as the resulting ranking of frontier productivity would be affected by the region underlying the comparison.

Nevertheless, the information it provides can be very useful for pair-wise comparisons of frontier productivity, leading to the identification of suitable benchmarks for a certain region. For example, the director in charge of the Northern region can explore the

⁴ It is 'superefficient' in relation to the spread observed in Lisbon, considered as the reference.

relative productivity of its regional frontier using the data reported in the last column of Table 8.3, and pictorially illustrated in Figure 8.4.

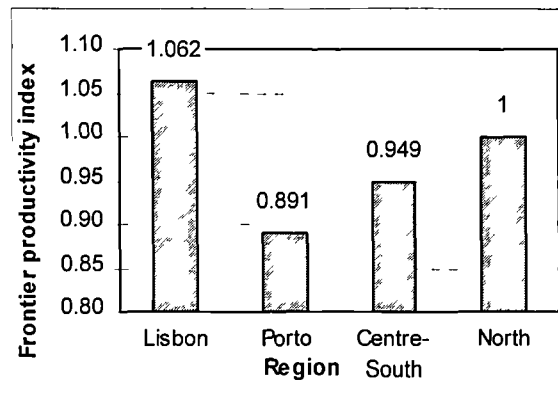


Figure 8.4 – Index for the comparison of frontier productivity, with the North as the reference region

From Figure 8.4 it can be concluded that the Northern region should direct a benchmarking effort to the branches in Lisbon, in order to identify the procedures that can improve the productivity of the best-practice branches in the North. Note that this type of analysis does not identify the *causes* of dominance of certain regional frontiers. In this example, the greater productivity observed in Lisbon may be due to the potential of the region in terms of socio-economic and demographic conditions, or to better managerial practices defined at the regional level, leading to greater productivity.

In order to obtain the productivity ranking of all regional frontiers, independent of the reference region underlying the comparison, the adjusted index $IF(\text{adj})_i^{\text{AB}}$, described in (8.24), should be used. Note that whilst the index for pair-wise comparisons (IF_i^{AB}) may be of interest to managers at the regional level, the adjusted index enabling an overall ranking of regional productivity is of interest to managers at a higher level, in charge of the entire network. The results of the adjusted index are shown in Table 8.4.

Table 8.4 - *Index for the ranking of frontier productivity of all regions* ($IF(adj)_i^{AB}$)

A \ B	Lisbon	Porto	Centre-South	North
Lisbon	1	0.877	0.937	0.978
Porto	1.141	1	1.069	1.115
Centre-South	1.067	0.935	1	1.043
North	1.023	0.897	0.959	1

The results in Table 8.4 are reported such that a value smaller than unity indicates that the region listed in the row heading has a frontier with greater productivity than the region listed in the column heading⁵.

Figure 8.5 pictorially illustrates the ranking of frontier productivity, using Lisbon as the reference (i.e., data taken from the first line in Table 8.4).

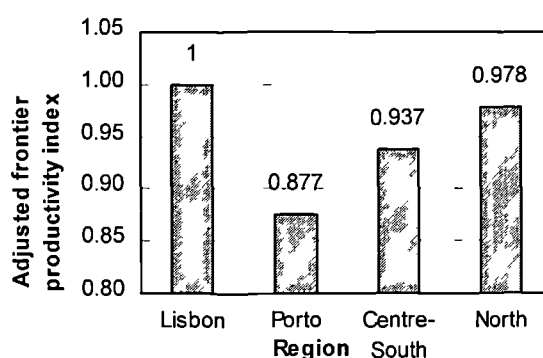


Figure 8.5 – *Adjusted index for the comparison of productivity of the regional best-practice frontiers, with Lisbon as the reference region*

Figure 8.5 shows that the frontier of Porto is considerably distant from the others. This indicates that the best-practice branches of Porto would be inefficient if compared to the best practice branches of the other regions. The frontier of Lisbon is the most productive, followed by the North and the Centre-South, respectively.

⁵ The circularity of this adjusted index can be verified: $IF(adj)_i^{LX,NT} \times IF(adj)_i^{NT,CS} = IF(adj)_i^{LX,CS}$, i.e., $0.9777 \times 0.9585 = 0.9371$.

Finally, the results of the overall regional comparison, using the index enabling an overall ranking of regional performance (i.e., $I(\text{adj})_i^{\text{AB}} = \text{IE}_i^{\text{AB}} \times \text{IF}(\text{adj})_i^{\text{AB}}$) are presented in Table 8.5.

Table 8.5 – Index for an overall comparison of regional performance ($I(\text{adj})_i^{\text{AB}}$)

A \ B	Lisbon	Porto	Centre-South	North
Lisbon	1	0.940	0.889	0.976
Porto	1.064	1	0.946	1.039
Centre-South	1.125	1.057	1	1.098
North	1.025	0.963	0.911	1

The results in Table 8.5 are reported such that a value smaller than unity indicates that the region listed in the row heading has a better performance status than the region listed in the column heading⁶.

Figure 8.6 pictorially illustrates these results using as reference the Lisbon region (i.e., the data used corresponds to the first line in Table 8.5).

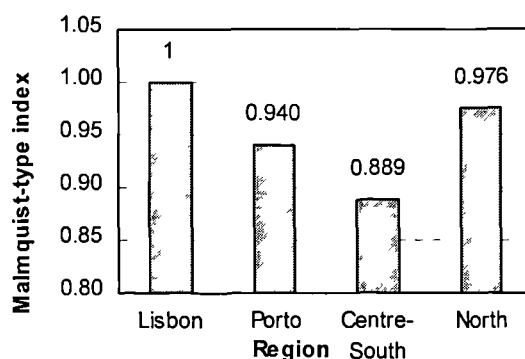


Figure 8.6 – Adjusted index for the ranking of regional performance

⁶ The circularity of this adjusted index can be verified: $I(\text{adj})_i^{\text{LX,NT}} \times I(\text{adj})_i^{\text{NT,CS}} = I(\text{adj})_i^{\text{LX,CS}}$, i.e., $0.976 \times 0.911 = 0.889$.

In terms of the managerial implication of the results obtained using this methodology, it was found that Porto has the frontier with the lowest productivity, although all its branches operate very close to the regional best-practice frontier. This may be interpreted as an indication that the targets set for these branches are relatively easy to achieve, and thus most branches operate close to the regional best-practice levels. The market potential of Porto should be carefully analysed to verify if these branches can be set more demanding targets, in order to close their productivity gap with respect to other regions.

The Centre-South region has a rather different performance status. Its frontier is closer to the most productive frontier, corresponding to Lisbon. However, there is a significant variability in the efficiency levels of the Centre-South branches, which indicates that the performance improvements should be directed to increase the homogeneity of within-region efficiency.

In relation to the North, improvements to within-region efficiency levels and frontier productivity are attainable, although it is close to the best performing levels in both performance dimensions.

Finally, Lisbon should focus on increasing the homogeneity of branches' efficiency levels. Its frontier is the most productive among the four regions considered.

In terms of the overall performance ranking between the four regions, based on the adjusted index $I(\text{adj})_i^{\text{AB}}$, it was found that Lisbon is the best performing region, followed by the North, Porto and the Centre-South, respectively.

8.8 Summary and conclusions

This chapter developed a new method for the comparison of performance of groups of DMUs operating under different conditions. Attributing the measured performance differences to its sources is a very important issue in any managerial context. This issue was addressed in this chapter by decomposing the performance index developed into a part comparing the efficiency spread among the DMUs in each group, and a part capturing the difference in frontier productivity between the groups.

The method developed consisted of an adaptation of the Malmquist index. Conventionally, the comparison underlying the construction of the Malmquist index refers to the same DMU at two different points in time, but in general any two DMUs can be considered. The index developed in this chapter changes the focus of the comparison from the individual DMUs to groups of DMUs, providing a summary measure of relative performance between groups of DMUs operating under different conditions at the same moment in time. The index described in this chapter used an input orientation, but its adaptation to output-oriented assessments is straightforward.

One of the advantages of the use of this performance index is that it does not require detailed data to characterise the environmental conditions associated with each group of DMUs, which can be difficult to obtain in 'real-world' applications. Also, it does not require a subjective aggregation of data on individual DMUs prior to the construction of the index comparing the groups' performance.

The method described also considered comparisons involving more than two groups of DMUs, by using one of the groups as the reference for the construction of the index. In this context, one desirable property of the index is its circularity, which ensures that the value obtained for the index is the same irrespectively of the reference considered. In

order to guarantee the circularity of the index, it was proposed that an adjusted index should be used whenever performance rankings of more than two groups of DMUs are of interest.

The applicability of the method was illustrated in the context of a comparison of regional performance of bank branches within a country. In terms of within-region efficiency spreads, the small dispersion of efficiency levels among the branches is likely to be a result of successful managerial approaches in the region. For the regions with large spreads, the performance may be improved by promoting the co-operation with network managers of other regions in order to transfer the benchmarking procedures between branches. Rotating branch staff across regions may also facilitate the implementation of the best practices in all branches.

In relation to the differences in frontier productivity, these may be attributable to the socio-economic and competitive conditions of the markets, as well as to the different managerial approaches and procedures defined at the regional level. This information regarding regional productivity may be extremely valuable for informing decisions regarding the location of new branches.

CHAPTER 9

Summary, conclusions and directions for further research

9.1 Summary and conclusions

This thesis developed new models and methods based on the DEA technique for performance measurement and improvement in multi-unit organisations in the financial sector. The developments presented in this thesis were grounded on the analysis of a case study from a commercial bank, which guarantees that the extensions to the performance measurement methods are driven from the actual needs of the organisations – i.e., *application driven theory*. An effort was made to ensure that the models and methods developed are generic and applicable to other types of for-profit organisations outside the financial services sector. The use of new managerial approaches to performance measurement and improvement may represent a substantial competitive advantage to any organisation operating in the highly competitive markets of the modern day world.

These new models and methods are components of a comprehensive framework for the assessment and improvement of performance. This framework was developed in Chapter 5 and detailed in the following chapters of the thesis. It can complement other approaches for planning and controlling performance frequently used in organisations. In particular, this thesis illustrated how profitability measures can be integrated with efficiency measures.

Each aspect of organisational efficiency was explored in detail in the following chapters of the thesis (Chapters 6 and 7), based on the analysis of a bank branch network. This involved the development of enhanced models for efficiency measurement and target setting at the DMU level.

In addition, an enhanced method for comparing the performance of groups of DMUs operating under different conditions was developed in Chapter 8. This method can disentangle managerial inefficiencies from those associated with the environmental characteristics within which the DMUs are required to operate.

The remainder of this section summarises the aims, the models and methods developed, the conclusions and the main contributions to the subject analysed in each chapter of the thesis.

Chapter 1 exposed the relevance of the topic chosen for this thesis. This introductory chapter justified the analysis of financial institutions' performance using the DEA method. The research objectives and thesis structure were also described.

Chapter 2 gave an outline of the measurement of efficiency. It described the historical evolution of frontier analysis and provided an overview of the main frontier methods. As DEA is the subject area of this thesis, the core of this chapter consisted of an introduction to DEA, including the theory underlying the representation of the technology of production and the efficiency frontier, the main DEA models, and a discussion of the recent developments in the DEA literature.

Chapter 3 reviewed the literature on banking efficiency measurement. The information gathered was used to guide the choice of the novel themes and questions addressed in this thesis.

Since no bank seems to rely exclusively on a single approach for performance assessment, Chapter 3 discussed the comparative advantages and limitations of ‘traditional’ performance measures (e.g., accounting ratios and profitability) versus frontier methods. It was concluded that DEA and profitability measures should be seen as complementary approaches for the analysis of financial institutions’ performance. This motivated the use of both measures in the empirical part of this thesis (see Chapter 5).

The review in Chapter 3 also focused on the definition of banks’ activity and the selection of appropriate inputs and outputs for an efficiency assessment. This was found to be a controversial topic in the literature of financial institutions’ performance. The approaches most frequently used in empirical studies are the ‘production’ and ‘intermediation’ approaches, which focus on the operational activity and the outcomes of financial intermediation, respectively. It was concluded that neither of these approaches is perfect, because neither fully captures the dual roles of financial institutions, associated with satisfying the customers requirements effectively (i.e., the aim of the operational activity) and generating maximal revenues from the cost levels incurred (i.e., the aim of financial intermediation activity). Therefore, the two approaches should be used in conjunction, whenever there is sufficient data to implement this type of research design, as was done in Chapter 5. Nevertheless, each of these approaches has a specific scope and the results obtained can be particularly useful for certain purposes. As the last

chapters of the thesis were only concerned with branches operational activity, only the model based on the 'production' approach was used.

Chapter 3 also summarised the aims, methodologies and conclusions of previous studies on banks' efficiency, with emphasis on those using the DEA method. The information gathered enabled the connection of the issues analysed in this thesis with previous research in the field of financial institutions' performance measurement. Furthermore, it enabled discussing the empirical results obtained in the case study analysed in this thesis in the light of the results from previous studies.

Chapter 4 described the commercial bank used as the case study for this thesis, with emphasis on the methods currently used to assess the branches' performance. It also introduced the context of the bank's business activity, providing an overview of the Portuguese financial services sector.

Chapter 5 developed a framework for the assessment of performance, integrating efficiency and profitability dimensions. It was shown that the efficiency measures are an important complement to profitability evaluations. The efficiency assessment can be used as a tool to quantify the potential profitability improvements, as well as identify the targets leading to the achievement of maximal profits.

Chapter 5 developed a DEA model to enable the identification of both input and output inefficiencies considering an objective of cost minimisation. This model departs from existing models for efficiency measurement, as the outputs are not considered fixed although they are still restricted such that the current total revenue is not exceeded. This model also differs from existing methods for the identification of the maximal profit

achievable at each DMU, as all profitability gains must be achieved through the rationalisation of costs.

In the context of bank branches' assessment, no increase to total revenue was allowed because there was no evidence that the market could support any expansion to business levels. Also, from the short-term perspective considered in this chapter, taking business from competitors is difficult, as customers generally have a loyal relation with their main bank. However, changes to the business mix were allowed because they can be achieved by emphasising the sales of certain types of financial products (e.g., customers can be persuaded to take a different mix of financial products, such as investing their savings in securities rather than having term deposits).

The results of this model used in the analysis of bank branches' performance showed that the major source of branches' intermediation inefficiency is related to the balance of financial products sold. The branches' total cost could be reduced significantly if different output mixes were adopted. This result suggests that the literature should devote greater attention to the output side of banking business.

Another issue addressed in Chapter 5 related to the definition of branches' inputs and outputs for the assessment of efficiency. The empirical study in this thesis adopted a research design based on the use of both the 'production' and 'intermediation' approaches, corresponding to the analysis of operational activity and the outcomes of financial intermediation. This was in line with the conclusions drawn from the review in Chapter 3. This analysis showed that these approaches are a powerful tool when applied jointly.

It was found that the efficiency spread of branches' operational activity is larger than the efficiency spread of the intermediation activity. This can be explained by the fact that relevant aspects of branches' intermediation activity are defined centrally, whereas the operational performance is mostly determined by local management and the quality of branch staff. These findings motivated the detailed assessment of operational efficiency in the final chapters of the thesis.

The results of the efficiency assessment showed that most branches are smaller than efficient scale from an operational perspective. However, closing these branches or consolidating with branches from other networks in case of mergers should be done with caution. Some of the scale inefficiencies detected in the operational activity may be associated with improved customer convenience, which has a positive impact in the intermediation business and network profitability, as suggested by the non-existence of scale inefficiencies in the intermediation activity. Thus, the removal of this type of scale inefficiencies may result in the loss of customers and revenues, which may more than offset the cost savings. The elimination of other inefficiency sources (i.e., pure technical inefficiency, input allocative inefficiency and price inefficiency) should be given pre-emptive priority.

Chapter 6 focused on the analysis of the effect of scale size on operational efficiency. As concluded from the literature review in Chapter 3, the returns to scale properties of financial institutions' activity are still a controversial issue in the banking literature. The empirical studies often found conflicting results, which motivated the exploratory analysis of this chapter.

In relation to the characterisation of branches' returns to scale, the main contribution of this chapter was the use of a method that can provide an objective characterisation of the returns to scale nature of all branches, including those not operating on the frontier of the PPS. This enables the identification of the best direction (i.e., input or output orientation) for the elimination of the pure technical inefficiency at each branch. In the case of this network, it was found that from the 90 branches with pure technical inefficiencies, the majority (73) should adopt an output orientation for the elimination of inefficiencies, and only 27 should adopt an input orientation.

The analysis of Chapter 6 also explored the consequences of adopting the same orientation for the elimination of pure technical inefficiency in all branches. A method to compare the relative magnitude of the two components of technical inefficiency, corresponding to pure technical and scale inefficiency, was developed and illustrated. It was found that with an input orientation a substantial amount of scale inefficiency would persist, whilst with an output orientation the resulting scale inefficiency would be much smaller, as the branches would generally be closer to the MPSS.

However, the generalisation of the scale efficiency and returns to scale results obtained in this case study, indicating that most scale inefficiencies are due to operating under increasing returns to scale, would require further evidence from other empirical studies using the same methodology. These results may be influenced to some extent by the context of the branches' activity and should not be generalised for branches in other settings.

Chapter 6 also focused on the development of target setting methods to improve efficiency, taking into account practical issues relating to the implementation of the DEA results.

One of the methods strengthens the benchmarking capacity of DEA by restricting the peers for a DMU under assessment according to a similarity criteria defined *a-priori* by the decision makers. In the analysis of this branch network, the similarity criteria used related to scale size. The use of this method ensures that the operating practices of the peers identified by DEA are easily transferable to the DMU under assessment. This target setting method leads to a gradual improvement effort, as each DMU is evaluated against an 'intermediate frontier' spanned from peers with similar characteristics. As a result, the achievement of the true best-practice frontier may require several phases.

The other target setting method developed avoids the problems associated with the existence of multiple optimal solutions to the DEA models. It enables choosing a unique scale size target between the alternative MPSS targets that may exist for each DMU, such that the changes required to the DMUs' profile in order to achieve maximal productivity can be minimised. The criteria used to choose the target scale size of each DMU consisted of adopting the smallest MPSS for DMUs with IRS and the largest MPSS for DMUs with DRS.

Chapter 7 focused on the analysis of cost efficiency considering different price scenarios, of either fixed prices (with prices known or with price uncertainty at the DMU level) or variable prices (where adjustments to price levels are at the DMs' discretion).

In relation to the fixed prices scenario, the main contribution of Chapter 7 was the development of a method for the estimation of a confidence interval for the CE measure

in situations of price uncertainty. The bounds of this interval are obtained from assessments in the light of the most favourable price scenario (“optimistic” perspective) and the least favourable price scenario (“pessimistic” perspective). The results obtained in this case study showed that the DEA method can provide robust estimates of cost efficiency even with price uncertainty, as the average amplitude of the confidence interval defined by the Optimistic CE and Pessimistic CE measures, based only on regional prices information, was very small (3%).

In relation to the variable prices scenario, it was proposed that an additional efficiency measure should be used, i.e., price efficiency. It captures the extent to which costs are increased due to paying prices above the minimal levels observed in other DMUs operating in the same market. The results of the empirical analysis showed that this type of inefficiency caused a significant increase in branches’ operational costs.

In order to become fully efficient under a variable prices scenario and produce with minimal cost, the DMUs may have to adjust both input and price levels. Two alternative sets of targets can be considered. One involves paying the minimal market prices in all DMUs, which may involve significant changes to the input mix of some DMUs. The alternative target setting method developed in this chapter suggests keeping the input mix of each DMU closer to the original values, which involves some adjustment to prices in order to reflect as closely as possible the marginal rate of substitution between the inputs, and still produce with the minimal cost.

The advantage of this second target setting method is that it avoids significant changes to branches’ TE input levels and current operating practices. Its major limitation is that some inputs may be set price targets below the minimal levels observed in the market, in

exchange for keeping the input targets close to the original TE input levels, as well as reflecting adequately the productivity of the remaining inputs by paying them prices above the minimal market level. Note that the relative magnitude of the price targets suggested using this method is within the relative price bounds observed in the market.

In relation to the use of this target setting method for the branches analysed, the input most often set price targets below minimal market levels was branch/account managers, whereas administrative/commercial staff and tellers were generally set price targets above the minimal market levels. Alternatively, by paying all inputs the minimal prices observed in the market, the most significant staff reductions would involve the branch/account managers.

Chapter 8 developed a method for the comparison of the performance of groups of DMUs operating under different conditions. A new performance index was developed (based on the structure of the Malmquist index), which can be decomposed into an index for the comparison of within-group efficiency spread, reflecting managerial efficiency, and an index comparing the productivity of the best-practice frontiers, which is determined by the context within which the DMUs are required to operate.

One of the advantages of the use of this performance index is that it does not require detailed data to characterise the environmental conditions associated with each group of DMUs, which may be difficult to obtain in many empirical studies, as happened in the case study analysed in this thesis. Also, it does not require a subjective aggregation of data on individual DMUs prior to the construction of the index reflecting the groups' performance.

The method developed in Chapter 8 also considered comparisons involving more than two groups of DMUs, which requires using one of the groups as the reference for the construction of the index. In this context, one desirable property of the index is its circularity, which ensures that the resulting ranking is independent of the group considered as a reference. Therefore, it was proposed that an adjusted index, satisfying the circularity property should be used, which can be used whenever performance rankings involving more than two groups of DMUs are of interest.

The applicability of the index developed was illustrated in the context of the comparison of bank branches' regional performance. In relation to the within-region efficiency spread index, a small dispersion in efficiency levels is likely to be a result of successful managerial approaches within the region. The region with the smallest efficiency spread was Porto. For regions with large spreads, as the Centre-South, their efficiency level can be improved by promoting co-operation with the regional managers in charge of other regions with smaller efficiency spreads. Rotating branch staff across the regions can also facilitate transferring the best-practice procedures to branches in different locations.

In relation to differences in frontier productivity, these may be attributable to the socio-economic and competitive conditions of the regional markets, as well as to the different managerial approaches, procedures and targets defined at the regional level. The least productive regional frontier was found to be Porto, whereas Lisbon has the most productive frontier. The market potential of Porto should be carefully analysed to verify if these branches could be set more demanding targets in order to close their productivity gap with respect to the other regions.

In summary, the main contributions of this thesis are:

- The development of a comprehensive framework for the analysis of performance in financial institutions, including the assessment of operational and financial intermediation efficiency and their impact on profitability.
- The development of a DEA model for the identification of inefficiencies in both input and output levels, considering an objective of cost minimisation, which should be used in situations where the market potential of each DMU is fully explored (such that revenue increases are not feasible).
- The use of a method that can objectively characterise the returns to scale nature of all DMUs, including those not operating on the frontier of the PPS. This method enables the identification of the best direction for the elimination of inefficiencies (i.e., input versus output orientations) in order to achieve high productivity levels.
- The development of a model that strengthens the benchmarking capacity of DEA by ensuring that the peers used in the efficiency assessment have similar characteristics to the DMU under analysis.
- The development of a method to choose a unique MPSS target for each DMU, such that the target scale size is as close as possible to the current scale size of the DMU. In case the DEA model has multiple optimal solutions, this involves the identification of the smallest MPSS for DMUs with IRS and the largest MPSS for DMUs with DRS.
- The extension of CE measurement to situations of price uncertainty at the DMU level. This thesis developed a new DEA model that can assess CE efficiency with a pessimistic perspective (in the light of the least favourable price scenario). The resulting Pessimistic CE measure, associated with an Optimistic CE measure, constitute the bounds of a confidence interval for estimating CE.
- The extension of CE measurement to situations where price adjustments are possible. This involved the definition of a new efficiency measure (price efficiency) and the development of target setting methods for achieving production at minimal cost.
- The development of a method for comparing the performance of groups of DMUs operating under different conditions. This thesis proposed the use of a new performance index, which can be decomposed into an index for the comparison of within-group efficiency spread and an index reflecting the frontier productivity. This index does not require subjective aggregations of data relating to individual DMUs. It can be adjusted to satisfy the circular test, which enables obtaining a robust rankings of performance of several groups of DMUs that is independent of the reference group underlying the comparison.

In relation to the impact of this research on the bank, the findings reported were well received by bank management. The results obtained from the DEA analysis were considered a powerful tool to complement both the EAS and SIM methods currently used to analyse branches' performance. The benchmarking properties of DEA, and the identification of peer branches within the network, were found particularly useful to set targets that are well adjusted to the profile of each branch. This can contribute to the acceptance of the results by branch staff. The analysis of inefficiency sources can be particularly informative for planning the directions of performance improvements that should be emphasised at the bank branch network. It was decided to start planning an implementation phase, by incorporating the DEA method in the banks' decision support tools. Overall, this thesis showed the usefulness of DEA as a tool to inform bank managers both with respect to the optimal strategies regarding the development of the branch network and to set targets to improve both efficiency and profitability levels.

9.2 Directions for further research

This thesis set out to analyse the performance of organisational units (bank branches) within the same institution (a commercial bank), which lead to an internal benchmarking effort. It would be interesting to extend this analysis by including branches from other networks, eventually from banks in different countries. This would broaden the scope of the performance improvement targets and promote learning from the best-practices observed in other organisations. Furthermore, with the financial markets of most countries becoming increasingly integrated, inter-country comparisons of banking institutions could provide valuable information regarding the consequences of this progressive integration. Banks that see the mutual benefits to be gained from exchanging information can overcome the problems of data confidentiality.

Another interesting research topic would be the comparison of the results obtained in this thesis, based on the DEA method, with those from alternative frontier methods, e.g., based on the use of parametric stochastic approaches for the estimation of the efficiency frontier. Further analysis using other frontier methods could provide additional validation of the results, and improve managerial confidence to address the issues raised by the assessment of branches' performance.

Quality of service in financial institutions is gaining paramount importance due to the increased competition in financial markets. Service quality is becoming a core element of differentiation required to attract new customers and retain the existing ones. In this context, another important research topic would be the analysis of the relation between branches' efficiency level and the quality of service provided to customers. A complete picture of branches' performance could involve the adaptation of the efficiency-profitability matrix used in Chapter 5 to also include a dimension relating to service quality.

Another research objective could be to shift the focus from the organisational level to the industry level. In particular, the Portuguese financial sector, where the bank used in the empirical part of this thesis operates, would be an interesting case study. Although its liberalisation process was one of the fastest in Europe – considering the starting point in terms of regulation and the pace of introduction of the changes towards the single European market of financial services – the impact of the changes on financial institutions' efficiency has not yet been analysed in depth.

It would also be important to monitor the evolution of financial institutions' performance over time. The analysis of efficiency and productivity change over time can be based on

the use of the Malmquist index, which relies on DEA efficiency measures. At the organisational level, management generally seeks to identify performance trends over time before taking important decisions that may affect the banks' results. This is essential to implement a successful performance improvement culture and ensure long-term viability. At the industry level, the analysis of performance trends over time could be used by the supervision and regulatory entities in each country to clarify the impact of deregulation in financial institutions' performance and to prevent bank's failure.

Appendix to Chapter 3

Summary of non-parametric studies on banking efficiency

Author and year	Inputs	Outputs	Orientation & Eff. Measures	Methods	Country and sample	Main Aims and Conclusions
Al-Faraj T.N., Alidi A.S., Bu-Bshait K.A. (1993)	No. of employees % Employees with college degree Average no. of years of experience Location index Highest authority rank index (%) Index for expenditure on decoration (%) Index for average monthly salaries (%) Index for other operational expenses (%)	Average monthly net profit Average monthly balance of current accounts Average monthly balance of savings accounts Average monthly value of other accounts Average monthly value of mortgages Index for loans (%) No. of current accounts	Input TE	DEA	15 bank branches from a commercial bank in Saudi Arabia	This study included a large number of inputs and outputs and only 3 branches were found inefficient.
Aly H.Y., Grabowski R., Pasurka C., Ragan N. (1990)	No. full-time employees Book value of premises and fixed assets (\$) Loanable funds (\$)	Real estate loans (\$) Commercial and industrial loans (\$) Consumer loans (\$) All other loans (\$) Demand deposits (\$)	Input CE AE TE PTE SE	DEA and Regression	1986 data on 322 US banks	First study to measure IAE of banks. It found that IAE was higher than TE, and SE higher than PTE. No significant efficiency differences between unit and branch organisational forms were found. Regression was used to analyse the effects of size, product diversity and urbanisation on efficiency.
Athanassopoulos A. (1998)	Market efficiency No. of transactions Potential market Sales representatives Internal automatic facilities No. branch outlets in the surrounding area Cost efficiency Direct labour costs Total technology facilities	Liability sales Loans and mortgages Insurances and securities Number of cards No. of transactions Liability sales Loans and mortgages Insurances and securities No. of cards	Output TE SE Input TE PTE SE	DEA and Multivariate statistical analysis	580 bank branches from a commercial bank in UK	This study used Pre-DEA clustering to ensure homogeneity among the branches assessed. Non-discretionary variables were included in the DEA model. The efficiency measure was separated into market and cost components. The cost efficiency component was found to be lower than the market efficiency component. The correlation between market and cost efficiencies was not significant. Few branches were operating at the MPSS from either a market or a cost perspective.
Athanassopoulos A. (1997)	Production approach No. of employees No. ATMs and teller machines No. of computers terminals Intermediation approach Non-interest costs Interest costs	No. of deposit accounts No. of credit transaction No. of debit transactions No. of loan applications evaluated No. transactions involving commissions Non-interest income (\$) Volume of loans (\$) Time deposit accounts (\$) Savings deposit accounts (\$) Current deposit accounts (\$)	Input TE Non-radial TE	DEA and Regression analysis	68 bank branches from a commercial bank in Greece	Branches efficiency was analysed under both the production and intermediation approaches. Weights restrictions were included in the production efficiency model to incorporate managerial perceptions. The intermediation approach explored the potential for both input reductions and output expansions. Regression analysis was used to explore the impact of service quality on branches' efficiency.

Author and year	Inputs	Outputs	Orientation & Eff. Measures	Methods	Country and sample	Main Aims and Conclusions
Barr R.S. Seiford L.M. Siems T. (1994)	Full-time equivalent employees Salary expenses (\$) Premises and fixed assets (\$) Other non-interest expenses (\$) Total interest expense (\$) Purchased funds (\$)	Core Deposits (\$) Earning assets (\$) Total interest income (\$)	Input TE	DEA and Regression	1984-1989 data on 930 US banks	This study found that the banks that survive can be differentiated from banks that fail using the DEA efficiency score. Two new bank failure prediction models were developed using Probit Regression. These models provided one-year-ahead and two-years-ahead predictions. The banks that survive were differentiated from the banks that fail using the DEA efficiency score. This difference could be statistically detected long before failure occurs.
Barr R.S. Seiford L.M. Siems T. (1993)	Full-time equivalent employees Salary expenses (\$) Premises and fixed assets (\$) Other non-interest expenses (\$) Total interest expense (\$) Purchased funds (\$)	Core Deposits (\$) Earning assets (\$) Total interest income (\$)	Input TE	DEA	1984-1989 data on 930 US banks	The banks that survive were differentiated from the banks that fail using the DEA efficiency score. This difference could be statistically detected long before failure occurs.
Berg S.A. Forsund F.R. Hjalmarsson L. Suominen M. (1993)	Man-hours per year Book value of machinery and equipment (\$)	Total loans (\$) Total deposits (\$) No. of branches	Input TE PTE SE	DEA and Malmquist index	1990 data on 503 Finnish 150 Norwegian 126 Swedish banks	First DEA study to compare banks' productivity across countries. The Malmquist index was adapted to allow a comparison between the performance of groups of DMUs at a given moment in time. The index developed was decomposed into an index reflecting the productivity of the countries' best practice frontiers and an index of within-country efficiency spread. The larger within-country efficiency spreads were found in Finland and Norway. Most of the overall best performing banks were Swedish.
Berg S.A. Forsund F.R. Jasen E.S. (1992)	Employees-hours per year Materials costs (i.e., operating costs divided by a materials price index)	Short-term loans (\$) Long-term loans (\$) Produced deposits (\$) (Negative) Loan losses (\$)	Input TE	DEA and Malmquist index	1980-1989 data on 152 Norwegian banks	This paper analysed productivity change during deregulation periods. It was found productivity regress at the average bank prior to deregulation, but rapid growth afterwards. Little productivity growth occurred due to the shift of the frontier. Instead, it was due to less dispersion in efficiency levels within the banking sector. The introduction of loan losses in the output vector as a measure of risk exposure stemming from poor loan evaluations caused only minor changes to the results.
Berg S.A. Forsund F.R. Jasen E.S. (1991)	First approach Labour costs Machines costs Materials costs Buildings costs Second approach Labour costs Machines costs Materials costs Buildings costs	Demand deposits Time deposits Short-term loans Long-term loans (measured by no. accounts & average size) Demand deposits (\$) Time deposits (\$) Short-term loans (\$) Long-term loans (\$)	Input TE PTE SE	DEA	1985 data on 107 Norwegian banks	This paper analysed the influence of output metric on the efficiency results. Output levels were measured by the no. of accounts and their average size, versus the total balances of the accounts. The average efficiency levels and the characterisation of the type of returns to scale occurring at each DMU were not significantly affected by the output specification used. However, the efficiency ranking of individual DMUs was heavily dependent on the output metric used.

Author and year	Inputs	Outputs	Orientation & Eff. Measures	Methods	Country and sample	Main Aims and Conclusions
Bergendahl G. (1995)	Cost of personnel Cost of material Cost of funding	Loans (\$) Floating rate deposits (\$) Fixed rate deposits (\$) Interbank placement (\$) Net margins (\$)	Input TE	DEA and Financial Ratio analysis	1987-1991 data on 5 German banks	This paper analysed the evolution of efficiency of the five largest banking groups in Germany during the period when Bancassurance was introduced. The results of DEA and ratio analysis were combined to evaluate each of the banks considered.
Bergendahl G. (1998)	Cost of personnel Cost of material Cost of funding	Loans (\$) Deposits (\$) Gross revenues (\$)	Input TE PTE	DEA	1992-1993 data on 48 Nordic banks	This paper analysed the efficiency of Nordic banks (from Denmark, Finland, Norway and Sweden) using window analysis. Three alternative "windows" were considered (i.e., "window 92", "window 93 and "window 92-93").
Bhattacharyya A. Lovell C.A.K. Sahay P. (1997)	Interest expenses Operating expenses	Advances (\$) Investments (\$) Deposits (\$)	Output PTE	DEA and Stochastic Frontier Regression	1986-1991 data on 70 Indian banks	This paper used DEA to calculate efficiency scores during the initial stages of deregulation in India. A stochastic frontier regression method was then used to explain the variation in calculated efficiencies in terms of time factors and ownership characteristics. The publicly-owned banks were the most efficient prior to deregulation, but the foreign owned banks were rising their efficiency levels under increased competition.
Brockett P.L. Charles A. Cooper W.W. Huang Z.M. Sun D.B. (1997)	Interest expense (\$) Non-interest expense (\$) Provision for loan losses (\$) Total deposits (\$)	Interest income (\$) Total non-interest income (\$) Allowances for loan losses (\$) Total loans (\$)	Input TE	DEA	1984-1985 data on 21 US banks	This paper used DEA for monitoring bank performance. Cone ratio envelopments were used to transform original data in order to reflect the performances of "excellent" banks identified <i>a-priori</i> . The analyses considered issues of "risk coverage" in the efficiency assessment, which can be of use to regulators.
Charles A. Cooper W.W. Huang Z.M. (1990)	Total operating expense (\$) Total non-interest expense (\$) Provision for loan losses (\$) Actual loan losses (\$)	Total operating income (\$) Total interest income (\$) Total non-interest income (\$) Total net loans (\$)	Input TE	DEA	1980-1985 data on 48 US banks	This paper developed a cone ratio DEA model and illustrated its use with a banking application. The input variables were defined to include indicators of risk in banking. Information regarding "excellent" bank performance identified <i>a-priori</i> was included in the DEA assessment.
Chen T.Y. (1998)	Staff Assets Interest expense	Loan services Portfolio investment Interest income Non-interest income	Input TE PTE SE	DEA	1996 data on 34 Taiwanese banks	This paper compared the efficiency of the publicly owned Taiwanese banks with that of the private banks created after the deregulation of the financial sector. The results revealed that the efficiency of the publicly owned banks is lower than that of the privately owned banks. The main source of inefficiency for public banks is of scale rather than technical nature.
De Borger B. Ferrier G.D. Kerstens K. (1998)	Total no. employees Occupancy and equipment costs Expenditure on materials	No. demand deposit accounts No. time deposit accounts No. real estate loans No. instalment loans No. commercial loans	Radial Input Output Graph Non-radial	FDH	1984 data on 575 US banks	The aim of this study was to compare the efficiency distributions and ranking of individual DMUs when alternative efficiency metrics and orientations are used. The empirical evidence based on US banking data revealed wide differences in efficiency scores and individual DMUs rankings across alternative measures.

Author and year	Inputs	Outputs	Orientation & Eff. Measures	Methods	Country and sample	Main Aims and Conclusions
Drake L. Howcroft B. (1994)	No. of interview rooms No. of ATMs Square meters of branch space Management grades Clerical grades Stationery costs	Till transactions Lending products Deposit products Automated transfers Clearing items Ancillary business Insurance business	Input TE PTE SE	DEA and Correlation	Data on 190 bank branches from a UK clearing bank	This study analysed the components of the TE measure. Most of the observed inefficiency was attributed to pure technical rather than scale effects. The DEA results were correlated with cost-income ratios and total lending. It was concluded that higher efficiency is associated with higher profits and smaller branch sizes.
Drake L. Weyman-Jones T.G. (1996)	No. full-time employees Book value of premises and fixed assets Retail funds and deposits (\$) Non-retail funds and deposits (\$)	Mortgage loans (\$) Other commercial assets (\$) Liquid excess holdings (in excess of prudential minimum)	Input CE AE TE PTE SE	DEA and Stochastic Frontiers	1988 data on 46 UK building societies	This study was based on DEA and SFA methods. The DEA results indicated that cost inefficiency was mainly due to allocative inefficiency, whereas technical inefficiency was mainly due to scale inefficiency. Both DEA and SFA found that the DMUs were operating close to CRS. Both methods provided extremely similar efficiency rankings
Drake L. Weyman-Jones T.G. (1992)	No. full-time and part-time employees Book value of premises and fixed assets Retail funds and deposits (\$) Wholesale funds and deposits (\$) Number of branches	Mortgage loans (\$) Other commercial assets (\$) Liquid excess holdings (in excess of prudential minimum)	Input TE PTE SE	DEA, Regression and Correlation	1988 data on 76 UK building societies	This study analysed building societies. It found that in most DMUs the inefficiency resulted from both scale and pure technical sources. Correlation and regression were used to investigate the relation of asset size and branching on efficiency levels. Most building societies, particularly the largest ones, were found to operate under DRS.
Elyasiani E. Mehdian S. (1995)	No. full-time employees Premises and fixed assets (\$) Large certificates, time and saving deposits (\$) Demand deposits (\$)	Investment (\$) Real Estate Loans (\$) Commercial and industrial loans (\$) Other loans (\$)	Input TE PTE SE	DEA and Statistical Tests	1979 & 1986 data on 300 US banks	This paper analysed the relation between efficiency and size in US banking during the pre and post-deregulation eras. Small banks achieved technological progress over the period 1979-86, whereas large banks displayed regress. Also, prior to deregulation small banks had less efficiency spread than large banks, whereas after deregulation both were equally efficient with respect to their own frontiers.
Elyasiani E. Mehdian S. (1990)	No. full-time employees Premises and fixed assets (\$) Large certificates, time and saving deposits (\$) Demand deposits (\$)	Investment (\$) Real Estate Loans (\$) Commercial and industrial loans (\$) Other loans (\$)	Input TE	DEA Regression and Statistical Tests	1980 & 1985 data on 191 US banks	This paper analysed the technological change in large US banks between 1980 to 1985. It was found technological improvement during this period, biased towards a more labour intensive mix.
Elyasiani E. Mehdian S. (1992)	No. full-time employees Premises and fixed assets (\$) Large certificates, time and saving deposits (\$) Demand deposits (\$)	Investment (\$) Real Estate Loans (\$) Commercial and industrial loans (\$) Other loans (\$)	Input CE AE TE PTE SE	DEA and Statistical Tests	1988 data on 80+80 US banks	This paper analysed the relationship between ownership (i.e., minority owned vs. non-minority owned banks) and efficiency. It was concluded that the two groups of banks operate with different production technologies. The efficiency spread of the banks with respect to their group-specific frontier was found to be similar.

Author and year	Inputs	Outputs	Orientation & Eff. Measures	Methods	Country and sample	Main Aims and Conclusions
Elyasiani E. Mehdian S. Rezvanian R. (1994)	No. full-time employees Premises and fixed assets (\$) Loanable funds (\$)	Securities (\$) Real Estate Loans (\$) Commercial and industrial loans (\$) Consumer loans (\$) Demand deposits (\$)	Input CE AE TE	DEA Regression and Ratio analysis	1983 & 1987 data on 203 US banks	This paper analysed the relationship between accounting ratios and DEA efficiency measures. Regression analysis revealed a significant positive association between financial performance (measured by accounting ratios) and efficiency.
Favero C.A. Papi L. (1995)	First approach No. full-time employees Book value premises and fixed assets (\$) Loanable funds (Current, saving & certificate deposits and interbank loans) Financial capital available for investment (\$) Second approach No. full-time employees Book value premises and fixed assets (\$) Certificate deposits and interbank loans (\$) Financial capital available for investment (\$) Total cost (labour, capital and materials)	Loans (\$) Investment in securities and bonds (\$) Non-interest income Current and saving deposits (\$) Loans (\$) Investment in securities and bonds (\$) Non-interest income Demand deposits (\$) Real Estate loans (\$) Instalment loans (\$) Commercial loans (\$)	Input TE PTE SE	DEA and Regression	1991 data on 174 Italian banks	This study identified both technical and scale inefficiency in Italian banks. It was found that the efficiency results were not too sensitive to the alternative definitions of inputs and outputs considered. Regression analysis was used to investigate the determinants of efficiency in the Italian banking sector. It was found higher efficiency levels for diversified banks engaged in non-traditional activities. Also, larger banks had higher efficiency. Banks in Southern Italy were the least efficient. No significant relationship could be found between efficiency and ownership or market structure.
Ferrier G.D. Grosskopf S. Hayes K.J. Yaisawarng S. (1993)			Input CE AE TE PTE SE	DEA Bootstrapping and Statistical Tests	1984 data on 468 US banks	This paper developed a new method for evaluating economies of diversification. Diseconomies of diversification and technical inefficiency were found to cause larger cost increases than scale inefficiency. Bootstrapping methods were used to obtain statistical tests of the significance of these results.
Ferrier G.D. Hirschberg J.G. (1997)	Total number of employees Value of furniture, equipment & real estate (\$) Consumer deposit accounts (\$) Commercial deposit accounts (\$) Industrial deposit accounts (\$)	Consumer, commercial and industrial loans (\$) Deposits at other financial institutions (\$) Investments (\$) No. of branches	Input TE	DEA and Bootstrapping	1986 data on 94 Italian banks	The aim of this paper was to show how bootstrapping can be used to obtain a sampling distribution of efficiency scores for individual DMUs. An illustrative example with Italian banking data was used. The bootstrapping provided confidence intervals and a measure of bias for the original DEA efficiency scores.
Ferrier G.D. Kerstens K. Vanden Eeckaut P. (1994)	Total number of employees Occupancy and equipment costs Expenditure on materials	No. demand deposit accounts No. time deposit accounts No. real estate loans No. instalment loans No. commercial loans	Input PTE non-radial	DEA and Statistical Tests	1984 data on 575 US banks	This paper illustrated empirically the effects of using alternative efficiency measures, radial and non-radial (e.g., Russell, Zieschang & Asymmetric Fare) on a DEA frontier. Large differences were found in the distributions of the different efficiency measures obtained. The correlation between the efficiency measures obtained for each DMU was low.
Ferrier G.D. Lovell C.A.K. (1990)	Total number of employees Occupancy, furniture and equipment costs Expenditure on materials	No. demand deposit accounts No. time deposit accounts No. real estate loans No. instalment loans No. commercial loans	Input CE AE TE SE	DEA and Stochastic Frontiers	1984 data on 575 US banks	This paper compared DEA and stochastic frontiers as alternative methods for measuring efficiency in the banking sector. Several environmental variables were considered. Both methods yielded very similar results regarding the identification of the returns to scale nature. However, the estimates of cost efficiency and its partition into technical and allocative components differed between the methods.

Author and year	Inputs	Outputs	Orientation & Eff. Measures	Methods	Country and sample	Main Aims and Conclusions
Field K. (1990)	Labour Equipment Offices Agencies	Financial services supplied to depositors Newly advanced mortgages Previously advanced mortgages	Input TE PTE SE Congestion Eff.	DEA and Correlation	1981 data on 71 UK building societies	This study found that the major source of building societies inefficiency was due to scale inefficiency, rather than pure technical or congestion inefficiency. Most building societies had decreasing returns to scale. It was found a significant negative relationship between technical efficiency and size of the building societies.
Fried H.O. Lovell C.A.K. (1994)	First approach No. full-time employees Other operating expenses (\$)	No. loan accounts No. deposit accounts Reciprocal of loan price index (%) Deposit price index (%) Loan variety index Deposit variety index	Output	FDH and Regression analysis	1990 data on 8,947 US credit unions	This paper described a two-stages performance analysis of US credit units, focusing on services provided to members. The first stage focused on efficiency measurement using the FDH method, followed by regression analysis to analyse the impact of environmental factors on efficiency. The second stage of the project focused on efficiency improvement. Field testing with managers resulted in substantial adjustments to the variables specification. The inefficiencies detected with the new variable specification were smaller, although the frequency of dominance by other peer DMUs has increased. Regression analysis indicated that larger credit unions tended to be less efficient.
	Second approach Total operating expenses (\$)	No. loan accounts No. deposit accounts Reciprocal of loan price index (%) Deposit price index (%) Service variety index Transactions indicator				
Fried H.O. Lovell C.A.K. Vanden Eeckaut P. (1993)	No. full-time equivalent employees Other operating expenses (\$)	No. loan accounts No. deposit accounts Reciprocal of loan price index (%) Deposit price index (%) Loan variety index Deposit variety index	Output	FDH and Regression analysis	1990 data on 8,947 US credit unions	This study evaluated the performance of US credit unions based on their ability to provide services to their members. It was found greater scope for improvement in the quantity dimension than in the price or variety of output dimensions. The environmental variables included in the regression analysis explained a small but statistically significant portion of the performance differences identified.
Fried H.O. Lovell C.A.K. Vanden Eeckaut P. (1995)	No. full-time equivalent employees Other operating expenses (\$)	No. loan accounts No. deposit accounts Reciprocal of loan price index Deposit price index Loan variety index Deposit variety index	Output	FDH and Regression analysis	1990 data on 8,947 US credit unions	This study evaluated the performance of US credit unions based on the FDH technique. It extended the earlier analysis by focusing on the managerial implications of the efficiency evaluation for individual DMUs. Regression analysis was used to determine the extent to which inefficiencies are associated with environmental conditions.
Fukuyama H. (1995)	No. full-time employees Book value premises and real estate (\$) Value of liabilities (\$)	Revenue from loans (\$) Revenue from investments (\$)	Input TE	DEA Malmquist index and Correlation	1989-1991 data on Japanese banks (155 in 1989, 154 in 1990, 153 in 1991)	This paper analysed the impact of the collapse of the 'bubble economy' in Japan on the efficiency and productivity of banks, using Malmquist indexes. The relation between productivity and organisational form and size was investigated. Banks specialised in long-term activities were found to be more efficient, whereas small banks exhibited greater productivity gains. The collapse of the bubble economy appeared to have affected more seriously the efficiency of banks in 1991.

Author and year	Inputs	Outputs	Orientation & Eff. Measures	Methods	Country and sample	Main Aims and Conclusions
Fukuyama H. (1996)	No. full-time employees Book value premises and real estate (\$) Value of liabilities (\$)	Revenue from loans (\$) Revenue from investments (\$)	Input TE PTE SE	DEA and Correlation	1992 data on 435 Japanese credit associations	This study analysed the efficiency of credit associations in Japan. It was found that the major source of inefficiency is pure technical and not scale inefficiency. Larger credit associations were found to be more efficient. Both input and output orientations were considered in the characterisation of returns to scale.
Fukuyama H. (1993)	No. full-time employees Book value premises and real estate (\$) Value of liabilities (\$)	Revenue from loans (\$) Revenue from other business activities (\$)	Input TE PTE SE	DEA and Statistical Tests	1990 data on 143 Japanese banks	This study found that the major source of inefficiency in Japanese banks is pure technical and not scale inefficiency. Most of the scale inefficient banks were operating under increasing returns to scale. The relation between organisational status, size and efficiency was analysed. It was found that city banks had higher TE and larger banks had higher SE.
Garbaccio R.F. Hermalin B.E. Wallace N.E. (1994)	Labour Physical capital Demand and time deposits Federal Home Loan Bank advances Equity capital	Output set 1 Total assets (\$) Output set 2 Operating income (\$) Output set 3 Mortgages Consumer and commercial lending Mortgage sales Mortgage servicing Other assets	Input CE AE TE	DEA and Varian-style algebraic methods	1987 data on 1360 US savings and loan institutions	This study analysed the efficiency of savings & loan institutions using two different methods. Both methods found a statistically significant difference in the efficiency levels of solvent and insolvent firms. The efficiency differences could be identified at least 2 years before failure. Three different output sets were defined, all providing an accurate picture of the DMUs efficiency. However, the technical efficiency measure based on one of these output sets appeared to be less accurate to predict failure, possibly because it did not account for input prices.
Giokas D. (1991)	No. person-hours worked Square meters of utilised branch space Operating costs (excluding labour costs)	DEA: Weighted no. deposit transactions Weighted no. credit transactions Weighted no. foreign receipts transactions Loglinear function: Total weighted no. of transactions	Input TE PTE SE	DEA, Loglinear function estimation and Correlation	1988 data on 17 bank branches from a Greek commercial bank	This paper compared the efficiency results obtained with DEA and the Loglinear function estimation. The efficiency results of the two methods did not exhibit significant differences. Both methods indicated the existence of increasing returns to scale for most branches.
Golany B. Storbeck J.E. (1999)	Total number of hours worked for tellers Total number of hours worked for non-tellers Floor space of the branch (in square feet) Direct mailing expense per customer Employment rate	Volume of loans Volume of deposits Average number of accounts per customer Customer satisfaction indicator	Output PTE	DEA and Statistical Tests	1992-1993 data on 182 US bank branches	This paper evaluated trends in branches efficiency over time. The paper also developed new DEA tools for budgeting and target-setting, and to evaluate differences in performance across groups of branches with difference characteristics. The presentation of the DEA results was tailored to the requirements of managers at different levels - e.g. branch, regional or corporate managers.

Author and year	Inputs	Outputs	Orientation & Eff. Measures	Methods	Country and sample	Main Aims and Conclusions
Grabowski R. Rangan N. Rezvanian R. (1993)	No. full-time employees Book value premises and fixed assets (\$) Loanable funds (\$)	Real Estate loans (\$) Commercial and industrial loans (\$) Consumer loans (\$) Demand deposits (\$) Investment securities (\$)	Input CE AE TE PTE SE	DEA and Statistical Tests	1989 US data on 522 multibank holding companies & 407 branch banks	This paper compared the efficiency of multibank holding companies (MBHC) and branch banking (BB) organisational forms. It was concluded that BB organisations are more efficient than MBHCs. These results suggest that the development of MBHCs may have been induced primarily by the banking regulations. The analysis of the impact of scale size on efficiency revealed that larger banks are generally more efficient.
Grifell-Tatje E. Lovell C.A.K. (1997)	No. of employees Materials costs (\$) Direct costs of buildings & depreciation (\$)	No. of loan accounts No. of checking accounts No. of savings accounts	Output PTE	DEA and Malmquist index	1986-1991 data on Spanish savings banks (unbalanced sample size: from 77 to 56)	This paper developed a new decomposition of productivity change into change in efficiency, magnitude of technical change and bias of technical change. The approach was illustrated with Spanish saving banks during a deregulation period. Declining productivity was identified, despite improvements in operational efficiency and the successful adaptation of output mix to environment.
Grifell-Tatje E. Lovell C.A.K. (1996)	No. of employees Materials costs (\$) Direct costs of buildings & depreciation (\$)	No. of loan accounts No. of checking accounts No. of savings accounts No. branch offices	Output TE	DEA Malmquist index and Statistical Tests	1986-1991 data on Spanish savings banks (unbalanced sample: 77 to 56)	This paper explored productivity change in Spanish savings banks during deregulation. It was found productivity decline, despite improvements in efficiency over time. The impact of branching and mergers on productivity was also analysed. It indicated that fast-branching banks experienced slower productivity decline than slow branching banks. No productivity gains following mergers and acquisitions were found during the period analysed.
Grifell-Tatje E. Lovell C.A.K. (1997)	No. of employees Costs of materials, buildings & depreciation (\$)	Value of loan accounts (\$) Value of checking accounts (\$) Value of savings accounts (\$)	Output TE PTE SE	DEA and Malmquist index	1986-1993 data on Spanish banks (sample varies from 77 to 50 ¹ savings banks and 61 to 67 commercial banks)	This paper analysed productivity change in Spanish savings and commercial banks. First, the productivity was analysed separately within each group. It was found productivity growth in both sectors, due to technological improvement as well as reduction in efficiency spread. Then, the two groups were merged to compare the two types of organisational structure. The analysis of the pooled sample indicated that savings banks were a more efficient form of organisation. The contradicting findings from previous studies of the Spanish financial sector were attributed to the different output specification used in this study.
Grifell-Tatje E. Lovell C.A.K. Pastor J.T. (1998)	No. of employees Costs of materials, buildings & depreciation (\$)	Value of loan accounts (\$) Value of checking accounts (\$) Value of savings accounts (\$)	Output PTE	DEA and Malmquist index	1986-1993 data on Spanish commercial banks (unbalanced sample: 61 to 67)	This paper developed a new quasi-Malmquist index which incorporates slacks on the efficiency measure. The comparison of the alternative Malmquist indexes was empirically illustrated using Spanish banking data. The two indexes identified the same productivity trend over time, although the non-radial efficiency measure was lower than the radial measure. It was argued that the new index should be preferred in the presence of large slacks and where a decomposition of the index is not required.

Author and year	Inputs	Outputs	Orientation & Eff. Measures	Methods	Country and sample	Main Aims and Conclusions
Hartman T. Storbeck J.E. (1996)	Interest costs (\$) Credit losses (\$) No. employees Value of deposits (\$) Equity capital Physical capital Federal home loan bank advances (\$)	Interest revenue (\$)	Input TE PTE SE	DEA	1984-1992 data on 12 Swedish banks	This study analysed the efficiency of Swedish banks during a deregulation period. The input-output set emphasized on loan operations. The method used was 'window analysis' with non-overlapping windows of three-year periods. The study found a decline of technical efficiency over time. The scale efficiency declined during this period despite the acquisitions involving some of the banks analysed.
Hermalin B.E. Wallace N.E. (1994)	No. employees Value of deposits (\$) Equity capital Physical capital Federal home loan bank advances (\$)	Total revenue	Input CE	DEA and Regression analysis	1986-1988 data on 1775 US Savings & Loan institutions	This study analysed the efficiency of Savings & Loan (S&L) institutions. Inefficient S&L were found more likely to fail in the future than efficient S&L. Controlling for lines of business emphasised, stock S&L were found more efficient than mutuals. Stock S&L were found more likely to focus on deregulated lines of business, positively correlated with insolvency.
Kantor J. Maital S. (1999)	Customer services assessment Labour costs Costs from services Floor space used for services (m ²) Bank transactions assessment Labour costs Costs from transactions Floor space used for transactions (m ²)	No. demand deposit accounts Customer services index Queue-replacing actions No. credit cards Transactions index Commissions received Savings accounts activities	Input TE	DEA ABC and Cluster Analysis	Data on 250 bank branches from a bank in the Middle East	This paper evaluated bank branches performance combining activity based accounting (ABC) and DEA. Two efficiency models were defined to enable a separate analysis of two types of branches' activities: providing services to customers and carrying out transactions. Cluster analysis was also used to enable benchmarking the branches' performance against a specific reference group, with common features with regard to business environment, mix of activities and size.
Lovell C.A.K. Pastor J.T. (1997)	No input	17 performance targets set by the bank (i.e., demand, high yield demand, time and home purchase deposits; personnel, credit card and mortgage loans; line-of-credit accounts, national commercial discounts, portfolio management, pension plans, investment funds, insurance policies; no. persons with direct deposits and credit cards, co-signed loans and reciprocal of delinquencies).	Input PTE	DEA	1995 data on 545 bank branches from a Spanish bank	This paper examined the performance of a target setting procedure used at branches from a Spanish bank. The study evaluated the ability of the branches to meet the targets set by bank management. The targets set for each branch were also evaluated. It was found that the list of variables for which targets are currently defined could be substantially reduced without significant loss or distortion of the information made available to bank management.
Miller S.M. Noulas A.G. (1996)	Transactions deposits (\$) Non-transactions deposits (\$) Interest expense (\$) Non-interest expense (\$)	Commercial and industrial loans (\$) Consumer loans (\$) Real estate loans (\$) Investments (\$) Interest income (\$) Non-interest income (\$)	Input TE PTE SE	DEA and Regression analysis	1984-1990 data on 201 US banks	This paper has analysed the efficiency of large US banks. The average technical efficiency found was very low. Regression analysis was then used to explore the effects of bank size, profitability, market power and location on PTE. Larger and more profitable banks were found to have higher PTE.

Author and year	Inputs	Outputs	Orientation & Eff. Measures	Methods	Country and sample	Main Aims and Conclusions
Oral M. Kettani O. Yolalan R. (1992)	Productivity assessment No. of personnel No. of on-line terminals No. of commercial accounts No. of saving accounts No. of checking accounts No. of credit applications	Amount of standard time spent on all kinds of transactions	Input TE	DEA and Statistical Tests	44 bank branches from a Turkish bank	This paper analysed the efficiency of an homogeneous set of bank branches from a Turkish bank. Two new formulations of DEA were explored, which forced certain bank branches to be part of the reference set of the branch under analysis. These new models were aimed to increase the managerial confidence on DEA results. Two alternative input-output sets were specified, focusing on branches' productivity and profitability. The results indicated that productivity and profitability rankings obtained with the alternative input-output specifications had a positive association.
	Profitability assessment Personnel costs (\$) Administrative expenses (\$) Depreciation (\$) Non-interest expenses (\$) Interests paid on deposits (\$)	Interest earned on loans (\$) Non-interest income (\$)				
	Productivity assessment No. of personnel No. of on-line terminals No. of commercial accounts No. of saving accounts No. of credit applications	Time spent on general service transactions Time spent on credit transactions Time spent on deposit transactions Time spent on foreign exchange transactions				
Oral M. Yolalan R. (1990)	Profitability assessment Personnel costs (\$) Administrative expenses (\$) Depreciation (\$) Interests paid on deposits (\$)	Interest earned on loans Non-interest income	Input TE	DEA	20 bank branches from a Turkish bank	This paper analysed the efficiency of an homogeneous set of bank branches from a Turkish bank. Alternative input-output combinations were considered in order to identify the one that best discriminates between efficient and inefficient branches. Two new formulations of DEA were explored, which forced certain bank branches to be part of the reference set of the branch under analysis. These new models were aimed to increase the managerial confidence on DEA results. It was found that more productive branches were generally more profitable as well.
	Productivity assessment No. of personnel No. of on-line terminals No. of commercial accounts No. of saving accounts No. of credit applications	Time spent on general service transactions Time spent on credit transactions Time spent on deposit transactions Time spent on foreign exchange transactions				
	Profitability assessment Personnel costs (\$) Administrative expenses (\$) Depreciation (\$) Interests paid on deposits (\$)	Interest earned on loans Non-interest income				
Parkan C. (1987)	No. full-time equivalent employees Annual rent (\$) Telephone/stationary expenses (\$) No. of on-line terminals Quality of customer service space ranking Marketing activity ranking	Basic transactions (weighted by standard time) Commercial account openings (index) Retail account openings (index) No. of loan applications Customer service survey rating No. corrections per no. transactions (inverse)	Output TE	DEA	35 bank branches from a Canadian bank	This paper analysed the efficiency of bank branches from a Canadian bank. The definition of input and output variables attempted to capture both quantitative and qualitative issues of branches' activity.
Pastor J.T. Perez F. Quesada J. (1997)	Non-interest costs (excluding personnel costs) Personnel costs (\$)	Value of loans (\$) Value other productive assets (\$) Value of deposits (\$)	Input TE	DEA and Malmquist index	1992 data on commercial banks from Spain (59), UK (18), Germany (22), France (67), Austria (45), Belgium (17), Italy (31), US (168)	The aim of this study was to analyse the productivity, efficiency and difference in technology of several banking systems. The Malmquist index was used for the comparison of productivity across countries at the same time period. Austria, Italy, Germany and Belgium were found to be the most productive countries, whereas USA, UK, France and Spain belonged to the less productive group. In terms of efficiency spread within countries, France, Spain and Belgium appeared as the most efficient banking systems.

Author and year	Inputs	Outputs	Orientation & Eff. Measures	Methods	Country and sample	Main Aims and Conclusions
Piesse J. Townsend R. (1995)	Profit objective Management costs Tangible fixed assets Number branches No. full-time equivalent employees Multiple objectives Management costs Tangible fixed assets Number branches No. full-time equivalent employees Interest paid on retail funds Interest paid on non-retail funds	Profit Interest earned from mortgage lending Interest earned from liquid assets No. depositors No. borrowers	Input TE PTE SE	DEA	1992 data on 57 UK building societies	This paper measured the efficiency of UK building societies. Five different input-output sets were specified, corresponding to alternative objectives. The first model used profit as the output variable, whereas the others segregated profits into interest earned on loans and interest paid on deposits. Some models included all the interest components whereas others omitted some interest components to reflect specific objectives of lenders or borrowers. As expected, the alternative variable specifications gave conflicting results. In terms of scale inefficiencies, DRS was found amongst larger societies.
Rangan N. Grabowski R. Aly H.Y. Pasurka C. (1988)	No. of full-time employees Book value of premises and fixed assets (\$) Purchased funds (\$)	Value real estate loans (\$) Value commercial and industrial loans (\$) Value consumer loans (\$) Value demand deposits (\$) Value time and saving deposits (\$)	Input TE PTE SE	DEA and Regression analysis	1986 data on 215 US banks	This paper analysed the efficiency of US banks. It was found that almost all the inefficiency detected was due to pure technical sources and not scale size effects. A second-stage regression analysis indicated that TE is positively related to size, negatively related to product diversity and not affected by the branching regulations.
Resti A. (1997)	No. of employees Adjusted book value of fixed capital (\$)	Value of deposits (\$) Value of loans (\$) Non-interest income (\$)	Input TE PTE	DEA FDH Stochastic Frontiers and Correlation	1988-1992 data on 270 Italian banks	This paper compared the efficiency estimates for Italian banks obtained from DEA and SFA. The results from both approaches were rather similar. In relation to the Italian banking sector, it was found a large efficiency spread, particularly in Southern banks. It was found a direct relationship between efficiency and asset quality. Also, the efficiency spread among Italian banks was not reduced over the period considered.
Schaffnit C. Rosen D. Paradi J.C. (1997)	No. tellers No. ledgers and accounting officers No. typing staff No. supervision personnel No. credit staff	No. counter transactions No. security sales No. deposit sales No. commercial loan sales No. personal loan sales No. term accounts No. commercial loan accounts No. personal loan accounts	Input TE PTE SE	DEA and Statistical Tests	1993 data on 291 Canadian bank branches	This paper focused on the efficiency of branches' personnel. Two alternative output specifications were considered, which included transactions or transactions and maintenance activities. The efficiency estimates were sharpened using input and output weights restrictions based on personnel salaries and standard times of services. The DEA results were also used for benchmarking purposes and to evaluate the branch groupings previously used at the bank for managerial purposes. Post-hoc statistical tests suggested that most efficient branches tended to be more profitable and deliver better quality of services. It was also identified a strong effect of branches' neighbourhood density on efficiency.

Author and year	Inputs	Outputs	Orientation & Eff. Measures	Methods	Country and sample	Main Aims and Conclusions
Sheldon G. (1994)	Bank deposits (\$) Total number of employees (\$) Office space and materials costs (\$)	Unsecured loans (\$) Non-real-estate-backed loans (\$) Real-estate-backed loans and mortgages (\$) Money market investments (\$) Securities (\$) Real estate investments (\$) Fee income (\$)	Input CE AE TE PTE SE	DEA and Stochastic Frontiers	1987-1991 data on 447 Swiss banks	This study analysed the efficiency of Swiss banks using both parametric and non-parametric techniques. Large cost inefficiencies were identified. Large banks were found to have higher PTE but smaller SE than the other banks. The analysis of economies of scope indicated that specialisation provides cost advantages. Productivity improved between 1987 and 1991 mainly due to the reduction of banks' efficiency spread.
Sheldon G. Haegler U. (1993)	Bank deposits (\$) Total number of employees (\$) Office space and materials costs (\$)	Unsecured loans (\$) Non-real-estate-backed loans (\$) Real-estate-backed loans and mortgages (\$) Money market investments (\$) Securities (\$) Real estate investments (\$) Fee income (\$)	Input CE AE TE PTE SE	DEA and Stochastic Frontiers	1987-1990 data on 477 Swiss banks	This study analysed the efficiency of Swiss banks using both parametric and non-parametric techniques. The efficiency results of the alternative model specifications were rather different, although both identified large cost inefficiencies. The main source of inefficiency appeared to be allocative. In relation to scale efficiency, the parametric results identified IRS in most branches, whereas the non-parametric results identified both IRS and DRS.
Sherman H.D. Gold F. (1985)	No. full-time equivalent employees (\$) Rent paid (\$) Total cost of supplies (\$)	No. more resource consuming transactions No. medium/high resource consuming trans. No. medium/low resource consuming trans. No. least resource consuming transactions	Input TE	DEA	Data on 14 bank branches from a US savings bank	This was the first paper to apply DEA to banking. This study set the pace for all subsequent studies on banking efficiency by showing the applicability of DEA. The output was measured by the number of transactions processed, which were grouped in four types according to complexity and level of resource consumption. The small sample used reduced the discriminatory power of DEA.
Sherman H.D. Ladino G. (1995)	No. full-time equivalent tellers No. full-time platform personnel No. full-time manager personnel Square feet of office space Operating cost (excluding personnel and rent)	No. deposits, withdrawals and checks cashed No. bond, bank & traveller checks transactions No. night deposits No. mortgage and consumer loans transactions No. new accounts Service quality index	Input TE	DEA	Data on 33 bank branches from a US bank	This paper reported an application of DEA to the analysis of efficiency of bank branches. The total potential savings identified were about \$9 million (US dollars). Actual changes that were implemented within the first year after the DEA study allowed the bank to save around \$6 million.
Soteriou A.C. Stavrinides Y. (1997)	No. hours worked by clerical personnel No. hours worked by managerial personnel No. computer terminal hours used Square meters of office space No. personnel accounts No. savings accounts No. business accounts No. credit application accounts	Foreign currency accounts Interbranch transactions Current and savings accounts Credit initialisations Loan renewals	Input Output TE	DEA	1994 data on 26 bank branches from a Cyprus bank	This paper focused on the improvement of service quality at bank branches. As noted by the researchers, the model cannot be used alone to assess performance, since it only considers a single service quality output and ignores other important outcomes of the branches' business. The model proposed in the paper must therefore be used in conjunction with other performance measures to evaluate the branches' activity.
Soteriou A.C. Zenios S.A. (1999)	Total branch costs	Foreign currency accounts Interbranch transactions Current and savings accounts Credit initialisations Loan renewals	Input CE SE	DEA	1994 data on 67 bank branches from a Cyprus bank	This paper developed an approach for estimating costs of financial products at the branch level. The novel approach described was based on the use of the DEA technique.

Author and year	Inputs	Outputs	Orientation & Eff. Measures	Methods	Country and sample	Main Aims and Conclusions
Taylor W.M. Thompson R.G. Thrall R.M. Dharmapala P.S. (1997)	Total deposits (\$) Total non-interest expense (\$)	Total (interest and non-interest) income (\$)	Input TE PTE	DEA	1989-1991 data on 13 Mexican commercial banks	This paper investigated the efficiency and profitability potential of Mexican banks. The efficiency was evaluated with weights restrictions (input-cone assurance regions) based on interest rates. This ensured that the weights underlying the DEA evaluation were economically reasonable. The banks' profitability potential was evaluated using profit ratio models, i.e., using weights restrictions defined as linked cone assurance regions.
Thompson R.G. Brinkmann E.J. Dharmapala P.S. Gonzalez-Lima M.D. Thrall R.M. (1997)	No. of employees Total deposits, borrowed funds & equity (\$)	Expected earnings (\$) (i.e., interest and non-interest earning less provisions for loan losses)	Input TE	DEA Factor analysis and Regression analysis	1986-1991 data on 100 large US banks	This paper analysed several issues relating to the evaluation of banks performance using DEA models. It discussed the efficiency measures obtained, the importance of slacks, and classified the branches into different efficiency groups. The stability of the efficiency evaluation to potential data errors was explored using sensitivity analysis. Assurance region bounds on weights were included in the analysis to sharpen the efficiency measure and identify the potential profit ratios. Statistical analysis highlighted the strengths of profit ratios to assess banks' performance.
Thompson R.G. Dharmapala P.S. Humphrey D.B. Taylor W.M. Thrall R.M. (1996)	No. of employees Book value of premises, furniture & equipment Total purchased funds (\$) Total no. of branches Total deposits (\$)	Total loans (\$) Total non-interest income (\$)	Input TE PTE	DEA	1980-1985 data on 48 large US commercial banks	This paper illustrated the computation of DEA efficiency and profit ratios, using assurance region bounds on the weights, with an application to large US banks. The weights bounds were based on price information. The maximal and minimal profit ratios were computed for the initial input-output levels of the banks.
Tulkens H. (1993)	Public Bank No. employees No. of windows operated No. automatic teller machines	No. checking & saving accounts transactions No. automatic teller machine transactions No. international transactions No. brokerage activities No. credit operations No. new accounts opened No. special services (e.g., card issues) Miscellany (e.g., insurance transactions)	Input TE PTE SE	FDH and DEA	1987 data on 773 & 804 bank branches of a public bank and 911 bank branches of a private bank in Belgium	This paper applied the FDH and DEA techniques to the assessment of bank branches. Firstly, the bank branches of a public bank were analysed both in terms of efficiency levels and dominance using the FDH technique. Then, the efficiencies of branches from the public and a private bank were compared using DEA and FDH. The outputs of the model and the no. of branches under analysis were slightly different from the first assessment. Using the DEA approach, the private bank appeared to perform somewhat better than the public bank. Conversely, the public bank scored better with the FDH model. It was considered that FDH provided a more reliable differentiation of branches' performance.
	Comparison of Private and Public Banks No. employees	No. transactions aggregated in 7 categories				

Author and year	Inputs	Outputs	Orientation & Eff. Measures	Methods	Country and sample	Main Aims and Conclusions
Vassiloglou M. Giokas D. (1990)	No. hours worked by personnel Costs of supplies Square meters of branch floor space No. of computer terminals	No. of "easiest" transactions No. of "medium-easy" transactions No. of "medium-difficult" transactions No. of "most difficult" transactions	Input TE	DEA	Data on 20 bank branches from a Greek bank	This paper studied the efficiency of a small set of bank branches all located in Athens. The outputs were measured by the no. of transactions and grouped in four classes according to their complexity, judged by responses of a sample of managers to a questionnaire. 11 out of the 20 branches were considered inefficient.
Weelock D.C. Wilson P.W. (1995)	Time and savings deposits (\$) Borrowed funds (\$) Book value of premises and fixed assets (\$) No. of bank officers (\$)	Total loans and bond holdings (\$) Demand deposits (\$)	Output TE	DEA and Statistical analysis	1910-1926 data on even numbered years for US banks	This paper used historical data to examine the causes of bank failure in periods of significant economic distress. The model of bank failure developed indicated that technically inefficient banks were more likely to fail than efficient banks.
Yeh Q. (1996)	Interest expenses (\$) Non-interest expenses (\$) Total deposits (\$)	Interest income (\$) Non-interest income (\$) Total loans (\$)	Input TE	DEA and Factor analysis	1981-1989 data on 6 commercial banks in Taiwan	This paper used DEA in conjunction with financial ratios to evaluate bank performance. Factor analysis was used to aggregate several ratios into meaningful dimensions that illustrate the financial strategies of the banks. The study concluded that the joint use of DEA and financial ratios can provide important insights relating to banks' performance.
Zaim O. (1995)	Total no. employees (\$) Total interest expenditures (\$) Depreciation expenditures (\$) Expenditures on materials (\$) Number of branches Institutional type (national or foreigner)	Total balance of demand deposits (\$) Total balance of time deposits (\$) Total balance of short-term loans (\$) Total balance of long-term loans (\$) Average size of demand deposit accounts (\$) Average size of time deposit accounts (\$)	Input CE AE PTE SE	DEA	1981 and 1990 data on Turkish banks (39 banks in 81 & 56 banks in 90)	This paper investigated the effects of liberalisation on the efficiency of Turkish banks. Two representative years for the pre and post liberalisation eras were chosen. The DEA models incorporated environmental variables in the input-output sets. It was found that the efficiency spread was reduced after liberalisation and that banks acquired a scale size closer to the optimal level. It was found that state banks were more efficient than private banks.
Zenios C.V. Zenios S.A. Agathocleous K. Soteriou A.C. (1999)	Managerial personnel Clerical personnel Computer terminals Space No. of current accounts No. of savings account No. foreign currency and commercial accounts No. credit applications	Total amount of work produced (measured in hours)	Input TE	DEA and Statistical Tests	Data on 145 bank branches from the Bank of Cyprus	This paper measured efficiency using an input set that captured the resources used and the no. of accounts held at each branch. These reflect the 'microenvironment' of the branch, achieved as a result of previous marketing efforts. The branches were separated in 3 groups (i.e., rural, urban and tourist branches) to explore the effect of environmental conditions on performance. The rural branches were found less efficient than urban branches. The performance of tourist branches is affected by seasonality. These branches were more efficient than urban branches only in the summer.

References for Appendix to Chapter 3

1. Al-Faraj T. N. , Alidi A. S. , and Bu-Bshait K. A. 1993. "Evaluation of Bank Branches by Means of Data Envelopment Analysis." *International Journal of Operations and Production Management* 13(9):45-52.
2. Aly H. Y. , Grabowski R. , Pasurka C. , and Rangan N. 1990. "Technical, Scale and Allocative Efficiencies in U.S. Banking: An Empirical Investigation." *Review of Economics and Statistics* 72:211-18.
3. Athanassopoulos A. D. 1998. "Nonparametric Frontier Models for Assessing the Market and Cost Efficiency of Large-Scale Bank Branch Networks." *Journal of Money, Credit and Banking* 30(2):172-92.
4. ———. 1997. "Service Quality and Operating Efficiency Synergies for Management Control in the Provision of Financial Services: Evidence From Greek Bank Branches." *European Journal of Operational Research* 98:300-313.
5. Barr R. S. , Seiford L. M., and Siems T. 1994. "Forecasting Bank Failure: A Non-Parametric Frontier Estimation Approach." *Recherches Economiques De Louvain* 60:417-29.
6. Barr R. S. , Seiford L. M. , and Siems T. F. 1993. "An Envelopment Analysis Approach to Measuring the Managerial Efficiency of Banks." *Annals of Operations Research* 45:1-19.
7. Berg S. A. , Forsund F. R. , Hjalmarsson L. , and Suominen M. 1993. "Banking Efficiency in the Nordic Countries." *Journal of Banking and Finance* 17:371-88.
8. Berg S. A. , Forsund F. R. , and Jansen E. S. 1992. "Malmquist Indices of Productivity Growth During the Deregulation of Norwegian Banking, 1980-89." *Scandinavian Journal of Economics* 94, supplement:S211-S228.
9. ———. 1991. "Technical Efficiency of Norwegian Banks: The Non-Parametric Approach to Efficiency Measurement." *Journal of Productivity Analysis* 2:127-42.
10. Bergendahl G. 1995. "Allfinanz, Bancassurance and the Future of Banking." *Applied Stochastic Models and Data Analysis* 11:323-47.
11. ———. 1998. "DEA and Benchmarks - an Application to Nordic Banks." *Annals of Operations Research* 82:233-49.
12. Bhattacharyya A. , Lovell C.A.K. , and Sahay P. 1997. "The Impact of Liberalisation on the Productive Efficiency of Indian Commercial Banks." *European Journal of Operational Research* 98:332-45.

13. Brockett P. L. , Charnes A. , Cooper W. W. , Huang Z. M. , and Sun D. B. 1997. "Data Transformations in DEA Cone Ratio Envelopment Approaches for Monitoring Bank Performances." *European Journal of Operational Research* 98:250-268.
14. Charnes A. , Cooper W. W. , Huang Z. M. , and Sun D. B. 1990. "Polyhedral Cone-Ratio DEA Models With an Illustrative Example to Large Commercial Banks." *Journal of Econometrics* 46:73-91.
15. Chen T. 1998. "A Study of Bank Efficiency and Ownership in Taiwan." *Applied Economics Letters* 5:613-16.
16. De Borger B. , Ferrier G. D. , and Kerstens K. 1998. "The Choice of a Technical Efficiency Measure on the Free Disposal Hull Technology: A Comparison Using US Banking Data." *European Journal of Operational Research* 105:427-46.
17. Drake L. and Howcroft B. 1994. "Relative Efficiency in the Branch Network of a U.K. Bank: An Empirical Study." *Omega, International Journal of Management Science* 22(1):83-90.
18. Drake L. and Weyman-Jones T. G. 1996. "Productive and Allocative Inefficiencies in U.K. Building Societies: a Comparison of Non-Parametric and Stochastic Frontier Techniques." *The Manchester School* 64(1):22-37.
19. ———. 1992. "Technical and Scale Efficiency in UK Building Societies." *Applied Financial Economics* 2(1-9).
20. Elyasiani E. and Mehdiian S. 1995. "The Comparative Efficiency Performance of Small and Large US Commercial Banks in the Pre- and Post-Deregulation Eras." *Applied Economics* 27:1069-79.
21. ———. 1990. "A Nonparametric Approach to Measurement of Efficiency and Technological Change: The Case of Large U.S. Commercial Banks." *Journal of Financial Services Research* 4(2):157-68.
22. ———. 1992. "Productive Efficiency Performance of Minority and Nonminority-Owned Banks: A Nonparametric Approach." *Journal of Banking and Finance* 16(5):993-48.
23. Elyasiani E. , Mehdiian S. , and Rezvanian R. 1994. "An Empirical Test of the Association Between Production and Financial Performance: the Case of the Commercial Banking Industry." *Applied Financial Economics* 4:55-59.
24. Favero C. A. and Papi L. 1995. "Technical Efficiency and Scale Efficiency in the Italian Banking Sector: a Non-Parametric Approach." *Applied Economics* 27:385-95.
25. Ferrier G. D. , Grosskopf S. , Hayes K. J. , and Yaisawarng S. 1993. "Economies of Diversification in the Banking Industry." *Journal of Monetary Economics* 31:229-49.

26. Ferrier G. D. and Hirschberg J. C. 1997. "Bootstrapping Confidence Intervals for Linear Programming Efficiency Scores: With an Illustration Using Italian Banking Data." *Journal of Productivity Analysis* 8:19-33.
27. Ferrier G. D. , Kerstens K. , and Vanden Eeckaut P. 1994. "Radial and Nonradial Technical Efficiency Measures on a DEA Reference Technology: A Comparison Using US Banking Data." *Recherches Economiques De Louvain* 60(4):449-79.
28. Ferrier G. D. and Lovell C. A. K. 1990. "Measuring Cost Efficiency in Banking: Econometric and Linear Programming Evidence." *Journal of Econometrics* 46:229-45.
29. Field K. 1990. "Production Efficiency of British Building Societies." *Applied Economics* 22:415-26.
30. Fried H.O. and Lovell C.A.K. 1994. "Enhancing the Performance of Credit Unions: The Evolution of a Methodology." *Recherches Economiques De Louvain* 60:431-47.
31. Fried H.O. , Lovell C.A.K. , and Vanden Eeckaut P. 1993. "Evaluating the Performance of US Credit Unions." *Journal of Banking and Finance* 17:251-65.
32. Fried H. O. , Lovell C. A. K. , and Vanden Eeckaut P. 1995. "Service Productivity in US Credit Unions." *The Service Productivity and Quality Challenge*, Harker P. T. Kluwer Academic Publishers.
33. Fukuyama H. 1995. "Measuring Efficiency and Productivity Growth in Japanese Banking: a Nonparametric Frontier Approach." *Applied Financial Economics* 5:95-107.
34. ———. 1996. "Returns to Scale and Efficiency of Credit Associations in Japan: A Nonparametric Frontier Approach." *Japan and the World Economy* 8:259-77.
35. ———. 1993. "Technical and Scale Efficiency of Japanese Commercial Banks: a Non-Parametric Approach." *Journal of Econometrics* 25:1101-12.
36. Garbaccio R.F. , Hermalin B.E. , and Wallace N.E. 1994. "A Comparison of Nonparametric Methods to Measure Efficiency in the Savings and Loan Industry." *Journal of the American Real Estate and Urban Economics Association* 22(1):169-93.
37. Giokas D. 1991. "Bank Branch Operating Efficiency: A Comparative Application of DEA and the Loglinear Model." *Omega, International Journal of Management Science* 19(6):549-57.
38. Golany B. and Storbeck J. 1999. "A Data Envelopment Analysis of the Operational Efficiency of Bank Branches." *Interfaces* 29(3):14-26.

39. Grabowski R. , Rangan N. , and Rezvanian R. 1993. "Organisational Forms in Banking: An Empirical Investigation of Cost Efficiency." *Journal of Banking and Finance* 17:531-38.
40. Grifell-Tatjé E. and Lovell C.A.K. 1997. "A DEA-Based Analysis of Productivity Change and Intertemporal Managerial Performance." *Annals of Operations Research* 73(177-189.).
41. Grifell-Tatjé E. and Lovell C. A. K. 1996. "Deregulation and Productivity Decline: The Case of Spanish Savings Banks." *European Economic Review* 40:1281-303.
42. ———. 1997. "The Sources of Productivity Change in Spanish Banking." *European Journal of Operational Research* 98:364-80.
43. Grifell-Tatjé E. , Lovell C.A.K. , and Pastor J. T. 1998. "A Quasi-Malmquist Productivity Index." *The Journal of Productivity Analysis* 10:7-20.
44. Hartman T. and Storbeck J. E. 1996. "Input Congestion in Loan Operations." *International Journal Production Economics* 46-47:413-21.
45. Hermalin B. E. and Wallace N. E. 1994. "The Determinants of Efficiency and Solvency in Savings and Loans." *RAND Journal of Economics* 25(3):361-81.
46. Kantor J. and Maital S. 1999. "Measuring Efficiency by Product Group: Integrating DEA With Activity-Based Accounting in a Large Mideast Bank." *Interfaces* 29(3):27-36.
47. Lovell C.A.K. and Pastor J.T. 1997. "Target Setting: An Application to a Bank Branch Network." *European Journal of Operational Research* 98:290-299.
48. Miller S. M. and Noulas A. G. 1996. "The Technical Efficiency of Large Bank Production." *Journal of Banking and Finance* 20:495-509.
49. Oral M. , Kettani O. , and Yolalan R. 1992. "An Empirical Study on Analysing the Productivity of Bank Branches." *IIE Transactions* 24(5):166-76.
50. Oral M. and Yolalan R. 1990. "An Empirical Study on Measuring Operating Efficiency and Profitability of Bank Branches." *European Journal of Operational Research* 46:282-94.
51. Parkan C. 1987. "Measuring the Efficiency of Service Operations: An Application to Bank Branches." *Engineering Costs and Production Economics* 12:237-42.
52. Pastor J. T. , Perez F. , and Quesada J. 1997. "Efficiency Analysis in Banking Firms: An International Comparison." *European Journal of Operational Research* 98:395-407.
53. Piesse J. and Townsend R. 1995. "The Measurement of Productive Efficiency in UK Building Societies." *Applied Financial Economics* 5:397-407.

54. Rangan N. , Grabowski R. , Aly H. Y. , and Pasurka C. 1988. "The Technical Efficiency of US Banks." *Economics Letters* 28:169-75.
55. Resti A. 1997. "Evaluating the Cost-Efficiency of the Italian Banking System: What Can Be Learned From the Joint Application of Parametric and Non-Parametric Techniques." *Journal of Banking and Finance* 21(2):221-50.
56. Schaffnit C. , Rosen D. , and Paradi J. C. 1997. "Best Practice Analysis of Bank Branches: An Application of DEA in a Large Canadian Bank." *European Journal of Operational Research* 98:269-89.
57. Sheldon G. 1994. "Economies, Inefficiencies and Technical Progress in Swiss Banking." Pp. 115-32 in *The Competitiveness of Financial Institutions and Centres in Europe*, vol. 28, Financial and Monetary Policy Studies, Fair D. and Raymond R. Dordrecht: Kluwer Academic Publishers.
58. Sheldon G. and Haegler U. 1993. "Economies of Scale and Scope and Inefficiencies in Swiss Banking." Pp. 103-40 in *Banking in Switzerland*, Blattner N. , Genberg H., and Swoboda A. (eds). New York: Springer-Verlag.
59. Sherman H. D. and Gold F. 1985. "Bank Branch Operating Efficiency: Evaluation With Data Envelopment Analysis." *Journal of Banking and Finance* 9:297-315.
60. Sherman H. D. and Ladino G. 1995. "Managing Bank Productivity Using Data Envelopment Analysis (DEA)." *Interfaces* 25(2):60-73.
61. Soteriou A.C. and Stavrinides Y. 1997. "An Internal Customer Service Quality Data Envelopment Analysis Model for Bank Branches." *International Journal of Operations and Production Management* 17(8):780-789.
62. Soteriou A.C. and Zenios S. A. 1999. "Using Data Envelopment Analysis for Costing Bank Products." *European Journal of Operational Research* 114:234-48.
63. Taylor W.M. , Thompson R. G. , Thrall R. M. , and Dharmapala P.S. 1997. "DEA/AR Efficiency and Profitability of Mexican Banks: A Total Income Model." *European Journal of Operational Research* 98:346-63.
64. Thompson R. G. , Brinkmann E. J. , Dharmapala P. S. , Gonzalez-Lima M. D. , and Thrall R. M. 1997. "DEA/AR Profit Ratios and Sensitivity of 100 Large U.S. Banks." *European Journal of Operational Research* 98:213-29.
65. Thompson R. G. , Dharmapala P. S. , Humphrey D. B. , Taylor W. M. , and Thrall R. M. 1996. "Computing DEA/AR Efficiency and Profit Ratio Measures With an Illustrative Bank Application." *Annals of Operations Research* 68:303-27.
66. Tulkens H. 1993. "On FDH Efficiency Analysis: Some Methodological Issues and Applications to Retail Banking, Courts and Urban Transit." *The Journal of Productivity Analysis* 4:183-210.

67. Vassiloglou M. and Giokas D. 1990. "A Study of the Relative Efficiency of Bank Branches: An Application of Data Envelopment Analysis." *Journal of the Operational Research Society* 41(7):591-97.
68. Wheelock D.C. and Wilson P. W. 1995. "Explaining Bank Failures: Deposit Insurance, Regulation, and Efficiency." *Review of Economics and Statistics* 77(4):689-700.
69. Yeh Q. 1996. "The Application of Data Envelopment Analysis in Conjunction With Financial Ratios for Bank Performance Evaluation." *Journal of the Operational Research Society* 47:980-988.
70. Zaim O. 1995. "The Effect of Financial Liberalization on the Efficiency of Turkish Commercial Banks." *Applied Financial Economics* 5:257-64.
71. Zenios C. V. , Zenios S. A. , Agathocleous K. , and Soteriou A.C. 1999. "Benchmarks of the Efficiency of Bank Branches." *Interfaces* 29(3):37-51.

Appendix to Chapter 7

Appendix 7A

Illustration of efficiency measurement in the weights space

This appendix details the calculations leading to the pictorial illustration of the technical and cost efficiency measures in the weights (price) space shown in Figure 7.3, based on the data set presented in Table 7.1, and reproduced below:

Table A.1 – *Inputs, output and prices at the DMUs*

	X_1	X_2	Y	P_1	P_2
DMU A	2	7	1	3	4
DMU B	3	5	1	4	3
DMU C	4	4	1	3	4
DMU D	5	3	1	4	3
DMU E	7	2	1	3	4
DMU F	3	7	1	3	4
DMU G	5	5	1	4	3
DMU H	6	3.6	1	4	3
DMU I	9	2	1	4	3
DMU J	10	2.5	1	3	4

The DEA model expressed in ratio form (i.e., model (2.6) in chapter 2) for the technical efficiency assessment of the DMUs considered is as follows:

DEA ratio model (for TE measurement)		(A7.1)
$\text{Max } \frac{y_{j_0} \mu}{x_{1j_0} v_1 + x_{2j_0} v_2}$		
s.t. A:	$\frac{1\mu}{2v_1 + 7v_2} \leq 1$	
B:	$\frac{1\mu}{3v_1 + 5v_2} \leq 1$	
C:	$\frac{1\mu}{4v_1 + 4v_2} \leq 1$	
D:	$\frac{1\mu}{5v_1 + 3v_2} \leq 1$	
E:	$\frac{1\mu}{7v_1 + 2v_2} \leq 1$	
F:	$\frac{1\mu}{3v_1 + 7v_2} \leq 1$	
G:	$\frac{1\mu}{5v_1 + 5v_2} \leq 1$	
H:	$\frac{1\mu}{6v_1 + 3.6v_2} \leq 1$	
I:	$\frac{1\mu}{9v_1 + 2v_2} \leq 1$	
J:	$\frac{1\mu}{10v_1 + 2.5v_2} \leq 1$	

In order to derive the efficiency scores, this model is transformed into a linear programming model. To facilitate its pictorial illustration, as all DMUs produce one unit of output, the normalisation required to obtain a linear programming model will be done on the output side, leading to an output oriented model. Note that under CRS the efficiency measure is the same irrespective of the model orientation. The normalisation

constant used is equal to the minimal cost of production of one unit of output (e.g. 27, observed in DMU B).

DEA model for TE assessment (in linear form)		(A7.2)
Min $\xi_{j_0} = x_{1j_0} v_1 + x_{2j_0} v_2$		
s.t.	$1 \cdot u = 27$	(A7.2a)
A:	$2v_1 + 7v_2 \geq u$	
B:	$3v_1 + 5v_2 \geq u$	
C:	$4v_1 + 4v_2 \geq u$	
D:	$5v_1 + 3v_2 \geq u$	
E:	$7v_1 + 2v_2 \geq u$	
F:	$3v_1 + 7v_2 \geq u$	
G:	$5v_1 + 5v_2 \geq u$	
H:	$6v_1 + 3.6v_2 \geq u$	
I:	$9v_1 + 2v_2 \geq u$	
J:	$10v_1 + 2.5v_2 \geq u$	

The pictorial representation of constraints A to J of model (A7.2) in Figure 7.3 considers that the right-hand-side is equal to 27 in all constraints, as imposed by the normalisation constant (A7.2a).

The corresponding technical efficiency measure for the DMU j_0 under assessment is

obtained as $\frac{1 \cdot u^*}{x_{1j_0} v_1^* + x_{2j_0} v_2^*}$ or $\frac{27}{\xi_{j_0}^*}$.

The measure of Farrell CE can be obtained solving a model with weights restrictions based on the DMUs' current prices. In the case of DMU F, the measure of Farrell CE can be obtained by adding to (A7.2) the following constraints, relating to the ratio of current prices (i.e., $P_1=3$ and $P_2=4$):

$$\frac{v_1}{v_2} = \frac{3}{4} \text{ or, in linear form, } 4v_1 - 3v_2 = 0$$

The measure of Optimistic CE is obtained by adding to (A7.2) the following input cone assurance region constraint:

$$\frac{3}{4} \leq \frac{v_1}{v_2} \leq \frac{4}{3} \text{ or, in linear form } \begin{cases} -4v_1 + 3v_2 \leq 0 \\ 3v_1 - 4v_2 \leq 0 \end{cases}$$

These restrictions change the focus of the assessment from technical efficiency towards cost efficiency measurement, assuming that only the maximal and minimal price bounds for all DMUs are known (i.e., $P_1^{\min} = P_2^{\min} = 3$ and $P_1^{\max} = P_2^{\max} = 4$)¹.

The calculation of Pessimistic CE requires the use of a different formulation of the efficiency model. For each DMU under assessment, the calculation of the Pessimistic CE measure involves solving a set of 10 different linear programming models. Each model considers a different DMU in the set under analysis as a potential peer. The Pessimistic CE value corresponds to the minimal efficiency value obtained at the optimal solution to the 10 models solved.

The model for the evaluation of Pessimistic CE for DMU F, using as peer DMU D, is illustrated below in (A7.3).

¹ The corresponding IC-AR is defined as $\left(\frac{P_1^{\min}}{P_2^{\max}} \leq \frac{v_1}{v_2} \leq \frac{P_1^{\max}}{P_2^{\min}} \right)$.

DEA model for Pessimistic CE assessment		(A7.3)
Max $\psi_{j_0j_D} = x_{1j_0} v_1 + x_{2j_0} v_2$		
s.t.	$1 \cdot u = 27$	
A:	$2v_1 + 7v_2 \geq u$	
B:	$3v_1 + 5v_2 \geq u$	
C:	$4v_1 + 4v_2 \geq u$	
D:	$5v_1 + 3v_2 = u$	
E:	$7v_1 + 2v_2 \geq u$	
F:	$3v_1 + 7v_2 \geq u$	
G:	$5v_1 + 5v_2 \geq u$	
H:	$6v_1 + 3.6v_2 \geq u$	
I:	$9v_1 + 2v_2 \geq u$	
J:	$10v_1 + 2.5v_2 \geq u$	
AR1:	$-4v_1 + 3v_2 \leq 0$	
AR2:	$3v_1 - 4v_2 \leq 0$	

The efficiency measure obtained by solving model (A7.3) is given by $\left(\frac{27}{\psi_{j_0j_D}^*} \right)$. The measure of Pessimistic CE for DMU F is obtained as the minimal among all the efficiency measures $\left(\frac{27}{\psi_{j_0j_P}^*} \right)$ corresponding to peers j_P ($j_P = A, \dots, J$) with feasible solutions to model (A7.3).

In this example, the peers associated with feasible solutions to the Pessimistic CE models solved for the assessment of DMU F would be DMUs B, C and D. The minimal value among the efficiency measures obtained with each of these three peers would correspond to an efficiency evaluation with DMU D as the peer.

Table A.2 reports the values of the TE, Farrell CE, Optimistic CE and Pessimistic CE for the 10 DMUs considered in this illustrative example.

Table A.2 – *Efficiency estimates for the example considered*

<i>DMU</i>	<i>TE</i>	<i>Optimistic CE</i>	<i>Farrell's CE</i>	<i>Pessimistic CE</i>
A	100%	93.1%	79.4%	79.4%
B	100%	100%	100%	93.1%
C	100%	100%	96.4%	96.4%
D	100%	100%	93.1%	93.1%
E	100%	93.1%	93.1%	79.4%
F	84.6%	81.8%	73.0%	73.0%
G	80%	80%	77.1%	77.1%
H	83.3%	83.3%	77.6%	77.6%
I	100%	77.1%	64.3%	64.3%
J	80%	67.5%	67.5%	56.8%

Appendix to Chapter 7

Appendix 7B

Illustration of the weights restrictions for relative prices used in the bank branches' assessment

This appendix shows the weights restrictions (defined as an input cone assurance region) used for the derivation of the 'implicit prices' and associated targets of branches in the Lisbon region, based on models (7.9) and (7.11) in Chapter 7. Note that the weights restrictions used to obtain the targets reported in Chapter 7 are regional-specific in order to reflect properly the market conditions of the branches' location.

Recall that an IC-AR is defined as:

$$\frac{P_{i^a}^{\min}}{P_{i^b}^{\max}} \leq \frac{v_{i^a}}{v_{i^b}} \leq \frac{P_{i^a}^{\max}}{P_{i^b}^{\min}}, \quad i^a < i^b, \quad i^a, i^b = 1, \dots, m, \quad (\text{B7.1})$$

where $P_{i^a}^{\min}$ and $P_{i^a}^{\max}$ stand for the minimal and maximal bounds estimated for the price of input i^a , respectively. $\frac{P_{i^a}^{\min}}{P_{i^b}^{\max}}$ and $\frac{P_{i^a}^{\max}}{P_{i^b}^{\min}}$ are the minimal and maximal bounds of the relative weights used in the assessment (v_{i^a} and v_{i^b} associated with inputs i^a and i^b).

The constraints in (B7.1) can be easily linearised in order to be included in the DEA linear programming model. The derivation of the two linear constraints relating to the relative value of the input weights for branch/account managers (input 1) and administrative/commercial staff (input 2) in the Lisbon region is explained next, based on information on the maximal and minimal regional prices reported in Table 7.5.

$$\frac{P_1^{\min}}{P_2^{\max}} \leq \frac{v_1}{v_2} \leq \frac{P_1^{\max}}{P_2^{\min}}$$

$$\Leftrightarrow \frac{4.56}{3.82} \leq \frac{v_1}{v_2} \leq \frac{6.21}{1.54}$$

$$\Leftrightarrow \begin{cases} -3.82 \cdot v_1 + 4.56 \cdot v_2 \leq 0 \\ 1.54 \cdot v_1 - 6.21 \cdot v_2 \leq 0 \end{cases}$$

For the branches in Lisbon region, the linear constraints for all input weights (i.e., prices) can be specified as follows¹, corresponding to the IC-AR restrictions in (7.9a) and (7.10a) in Chapter 7:

$$\begin{bmatrix} v1 & v2 & v3 & v4 \\ -3.82 & 4.56 & 0 & 0 \\ 1.54 & -6.21 & 0 & 0 \\ -5.16 & 0 & 4.56 & 0 \\ 3.78 & 0 & -6.21 & 0 \\ -1 & 0 & 0 & 4.56 \\ 1 & 0 & 0 & -6.21 \\ 0 & -5.16 & 1.54 & 0 \\ 0 & 3.78 & -3.82 & 0 \\ 0 & -1 & 0 & 1.54 \\ 0 & 1 & 0 & -3.82 \\ 0 & 0 & -1 & 3.78 \\ 0 & 0 & 1 & -5.16 \end{bmatrix} \begin{bmatrix} v1 \\ v2 \\ v3 \\ v4 \end{bmatrix} \leq 0$$

The parameters z_{ci} (used in models (7.10) and (7.11)) are illustrated for the first two lines of the matrix above, as follows:

$$\begin{bmatrix} z_{11} & z_{12} & z_{13} & z_{14} \\ z_{21} & z_{22} & z_{23} & z_{24} \\ \dots & \dots & \dots & \dots \end{bmatrix} = \begin{bmatrix} -3.82 & 4.56 & 0 & 0 \\ 1.54 & -6.21 & 0 & 0 \\ \dots & \dots & \dots & \dots \end{bmatrix}$$

¹ Note that if the DMUs evaluation is based on m inputs (4 in the DEA model defined for bank branches assessment), it is possible to generate C_2^m ($C_2^4 = 6$) input ratios, which give a total of 12 ($2 \times C_2^4$) linear inequality constraints for the IC-AR.

REFERENCES

1. Aigner D. J. and Chu S. F. 1968. "On Estimating the Industry Production Function." *American Economic Review* 58:826-39.
2. Aigner D. J. , Lovell C. A. K. , and Schmidt P. 1977. "Formulation and Estimation of Stochastic Frontier Production Models." *Journal of Econometrics* 6 (1):21-37.
3. Akhavein J.D. , Berger A.N. , and Humphrey D.B. 1997. "The Effects of Megamergers on Efficiency and Prices: Evidence From a Bank Profit Function." *Review of Industrial Organisation* 12(1):95-139.
4. Ali A. I. 1990. "Data Envelopment Analysis: Computational Issues." *Computers, Environment and Urban Systems* 14:157-65.
5. Allen L. and Rai A. 1996. "Operational Efficiency in Banking: An International Comparison." *Journal of Banking and Finance* 20:655-72.
6. Allen R. , Athanassopoulos A. D. , Dyson R. G. , and Thanassoulis E. 1997. "Weights Restrictions and Value Judgements in Data Envelopment Analysis: Evolution, Development and Future Directions." *Annals of Operations Research* 73:13-34.
7. Altunbas Y. , Molyneux P. , and Thornton J. 1997. "Big-Bank Mergers in Europe: An Analysis of the Cost Implications." *Economica* 64:317-29.
8. Aly H. Y. , Grabowski R. , Pasurka C. , and Rangan N. 1990. "Technical, Scale and Allocative Efficiencies in U.S. Banking: An Empirical Investigation." *Review of Economics and Statistics* 72:211-18.
9. Athanassopoulos, A. D. 1994. "Goal Programming and Data Envelopment Analysis (GoDEA) for Target-Based Multi-Level Planning: Allocating Central Grants to the Greek Local Authorities." *European Journal of Operational Research* 87:535-50.
10. Athanassopoulos A. D. 1998. "Nonparametric Frontier Models for Assessing the Market and Cost Efficiency of Large-Scale Bank Branch Networks." *Journal of Money, Credit and Banking* 30(2):172-92.
11. Athanassopoulos, A. D. 1995. "Performance Improvement Decision Aid Systems (PIDAS) in Retailing Organisations Using Data Envelopment Analysis." *Journal of Productivity Analysis* 6:153-70.
12. Athanassopoulos A. D. 1997. "Service Quality and Operating Efficiency Synergies for Management Control in the Provision of Financial Services: Evidence From Greek Bank Branches." *European Journal of Operational Research* 98:300-313.

13. Banker R. D. 1984. "Estimating Most Productive Scale Size Using Data Envelopment Analysis." *European Journal of Operational Research* 17:35-44.
14. Banker R.D. 1996. "Hypothesis Tests Using Data Envelopment Analysis." *Journal of Productivity Analysis* 7(2-3):139-59.
15. Banker R. D. 1993. "Maximum Likelihood, Consistency and Data Envelopment Analysis: A Statistical Foundation." *Management Science* 39(10):1265-73.
16. Banker R. D. , Bardhan I. R. , and Cooper W. W. 1996. "A Note on Returns to Scale in DEA." *European Journal of Operational Research* 88:583-85.
17. Banker R. D. , Chang H. , and Cooper W. W. 1996. "Equivalence and Implementation of Alternative Methods for Determining the Returns to Scale in Data Envelopment Analysis." *European Journal of Operational Research* 89(473-481).
18. Banker R.D. , Chang H. , and Cooper W.W. 1996. "Simulation Studies of Efficiency, Returns to Scale and Misspecification With Nonlinear Functions in DEA." *Annals of Operations Research* 66:233-53.
19. Banker R. D. , Charnes A. , and Cooper W. W. 1984. "Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis." *Management Science* 30(9):1078-92.
20. Banker R.D. , Charnes A. , Cooper W.W. , and Maindiratta A. 1988. "A Comparison of DEA and Translog Estimates of Production Frontiers Using Simulated Observations From a Known Technology." *Applications of Modern Production Theory: Efficiency and Productivity*, Dogramaci A. and Fare R. Kluwer Academic Publishers.
21. Banker R. D. , Charnes A. , Cooper W. W. , Swarts J. , and Thomas D. A. 1989. "An Introduction to Data Envelopment Analysis With Some of Its Models and Their Uses." *Research in Governmental and Nonprofit Accounting* 5:125-63.
22. Banker R. D. , Conrad R. F. , and Strauss R. P. 1986. "A Comparative Application of Data Envelopment Analysis and Translog Methods: An Illustrative Study of Hospital Production." *Management Science* 32(1):30-44.
23. Banker R.D. and Cooper W.W. 1994. "Validation and Generalization of DEA and Its Uses." *TOP - Sociedad Espanola De Estadistica e Investigacion Operativa* 2(2):249-314.
24. Banker R. D. , Gadh V. M. , and Gorr W. L. 1993. "A Monte Carlo Comparison of Two Production Frontier Estimation Methods: Corrected Ordinary Least Squares and Data Envelopment Analysis." *European Journal of Operational Research* 67(3):332-43.
25. Banker R. D. and Morey R. C. 1986. "Efficiency Analysis for Exogenously Fixed Inputs and Outputs." *Operations Research* 34(4):513-21.

26. ———. 1986. "The Use of Categorical Variables in Data Envelopment Analysis." *Management Science* 32(12):1613-27.
27. Banker R. D. and Thrall R. M. 1992. "Estimation of Returns to Scale Using Data Envelopment Analysis." *European Journal of Operational Research* 62:74-84.
28. Barr R. S. , Seiford L. M., and Siems T. 1994. "Forecasting Bank Failure: A Non-Parametric Frontier Estimation Approach." *Recherches Economiques De Louvain* 60:417-29.
29. Barr R. S. , Seiford L. M. , and Siems T. F. 1993. "An Envelopment Analysis Approach to Measuring the Managerial Efficiency of Banks." *Annals of Operations Research* 45:1-19.
30. Bauer P. W. 1990. "Recent Developments in the Econometric Estimation of Frontiers." *Journal of Econometrics* 46:39-56.
31. Belton V. and Vickers S.P. 1993. "Demystifying DEA - A Visual Interactive Approach Based on Multiple Criteria Analysis." *Journal of Operational Research Society* 44(9):883-96.
32. Benston G. J. 1965. "Branch Banking and Economies of Scale." *Journal of Finance* 20:312-31.
33. Benston G.J. , Hanweck G.A. , and Humphrey D.B. 1982. "Scale Economies in Banking: A Restructuring and Reassessment." *Journal of Money, Credit and Banking* 14(2):435-56.
34. Benston G. J. , Hunter W. C. , and Wall L. D. 1995. "Motivations for Bank Mergers and Acquisitions: Enhancing the Deposit Insurance Put Option Versus Earnings Diversifications." *Journal of Money, Credit and Banking* 27(3).
35. Berg S. A. , Forsund F. R. , Hjalmarsson L. , and Suominen M. 1993. "Banking Efficiency in the Nordic Countries." *Journal of Banking and Finance* 17:371-88.
36. Berg S. A. , Forsund F. R. , and Jansen E. S. 1992. "Malmquist Indices of Productivity Growth During the Deregulation of Norwegian Banking, 1980-89." *Scandinavian Journal of Economics* 94, supplement:S211-S228.
37. ———. 1991. "Technical Efficiency of Norwegian Banks: The Non-Parametric Approach to Efficiency Measurement." *Journal of Productivity Analysis* 2:127-42.
38. Bergendahl G. 1998. "DEA and Benchmarks - an Application to Nordic Banks." *Annals of Operations Research* 82:233-49.
39. Berger A.N. 1993. "Distribution-Free Estimates of Efficiency in the U.S. Banking Industry and Tests of the Standard Distributional Assumptions." *The Journal of Productivity Analysis* 4:261-92.

40. Berger A. N. , Hancock D. , and Humphrey D. B. 1993. "Bank Efficiency Derived From the Profit Function." *Journal of Banking and Finance* 17(2-3):317-47.
41. Berger A. N. , Hanweck G. A. , and Humphrey D. B. 1987. "Competitive Viability in Banking: Scale, Scope, and Product Mix Economies." *Journal of Monetary Economics* 20:501-20.
42. Berger A. N. and Humphrey D. B. 1991. "The Dominance of Inefficiencies Over Scale and Product Mix Economies in Banking." *Journal of Monetary Economics* 28:117-48.
43. ———. 1997. "Efficiency of Financial Institutions: International Survey and Directions for Future Research." *European Journal of Operational Research* 98:175-212.
44. ———. 1992. "Measurement and Efficiency Issues in Commercial Banking." Pp. 245-87 in *Output Measurement in the Service Sector*, Griliches Z. The University of Chicago Press.
45. Berger A.N. and Humphrey D.B. 1992. "Megamergers in Banking and the Use of Cost Efficiency As an Antitrust Defense." *Antitrust Bulletin* 37:541-600.
46. Berger A. N. , Hunter W. C. , and Timme S. G. 1993. "The Efficiency of Financial Institutions: a Review and Preview of Research Past, Present and Future." *Journal of Banking and Finance* 17(2-3):221-49.
47. Berger A. N. , Leusner J. H. , and Mingo J. J. 1997. "The Efficiency of Bank Branches." *Journal of Monetary Economics* 40:141-62.
48. Bhattacharyya A. , Lovell C.A.K. , and Sahay P. 1997. "The Impact of Liberalisation on the Productive Efficiency of Indian Commercial Banks." *European Journal of Operational Research* 98:332-45.
49. Borges A.M. 1993. "Portuguese Banking in the Single European Market." Pp. 325-43 in *European Banking in the 1990s* , ed. Dermine J. Great Britain: Blackwell Business.
50. Boufounou P. V. 1995. "Evaluating Bank Branch Location and Performance: A Case Study." *European Journal of Operational Research* 87:389-402.
51. Boussofiane A. , Dyson R. G. , and Thanassoulis E. 1991. "Applied Data Envelopment Analysis." *European Journal of Operational Research* 52:1-15.
52. Brealey RA and Myers SC. 1996. *Principles of Corporate Finance*. USA: McGraw-Hill (5th edition).
53. Brockett P. L. , Charnes A. , Cooper W. W. , Huang Z. M. , and Sun D. B. 1997. "Data Transformations in DEA Cone Ratio Envelopment Approaches for Monitoring Bank Performances." *European Journal of Operational Research* 98:250-268.

54. Canals J. 1993. *Competitive Strategies in European Banking*. United states: Oxford University Press.
55. ———. *Universal Banking*. Oxford: Clarendon Press.
56. Caves D. W. , Christensen L. R. , and Diewert W. E. 1982. "The Economic Theory of Index Numbers and the Measurement of Input, Output and Productivity." *Econometrica* 50(6):1393-414.
57. Chambers R. G. , Chung Y. , and Fare R. 1996. "Benefit and Distance Functions." *Journal of Economic Theory* 70:407-19.
58. Charnes A. , Clark C. T. , Cooper W. W. , and Golany B. 1985. "A Developmental Study of Data Envelopment Analysis in Measuring the Efficiency of Maintenance Units in the U.S. Air Forces." *Annals of Operations Research* 2:95-112.
59. Charnes A. and Cooper W. W. 1985. "Preface to Topics in Data Envelopment Analysis." *Annals of Operations Research* 2:59-94.
60. Charnes A. , Cooper W. W. , Huang Z. M. , and Sun D. B. 1990. "Polyhedral Cone-Ratio DEA Models With an Illustrative Example to Large Commercial Banks." *Journal of Econometrics* 46:73-91.
61. Charnes A. , Cooper W. W. , Lewin A. Y. , and Seiford L. M. 1994. *Data Envelopment Analysis: Theory, Methodology and Applications*. Klumer academic publishers.
62. Charnes A. , Cooper W. W. , and Rhodes E. 1981. "Evaluating Program and Managerial Efficiency: an Application of Data Envelopment Analysis to Program Follow Through." *Management Science* 27(6):668-97.
63. ———. 1978. "Measuring the Efficiency of Decision Making Units." *European Journal of Operational Research* 2: 429-44.
64. Chen T. 1998. "A Study of Bank Efficiency and Ownership in Taiwan." *Applied Economics Letters* 5:613-16.
65. Clawson, C. J. 1974. "Fitting Branch Locations, Performance Standards, and Marketing Strategies to Local Conditions." *Journal of Marketing* 38:8-14.
66. Cobb C. and Douglas P. 1928. "A Theory of Production." *American Economic Review* Supplement to Vol. 18:139-65.
67. Cook W. D. , Kress M. , and Seiford L. M. 1996. "Data Envelopment Analysis in the Presence of Both Quantitative and Qualitative Factors." *Journal of the Operational Research Society* 47(7):945-53.
68. ———. 1993. "On the Use of Ordinal Data in Data Envelopment Analysis." *Journal of the Operational Research Society* 44(2):133-40.

69. Cooper W.W. , Park K.S. , and Pastor J. T. 1999. "RAM: A Range Adjusted Measure of Inefficiency for Use With Additive Models, and Relations to Other Models and Measures in DEA." *The Journal of Productivity Analysis* 11:5-42.
70. Cooper W.W. , Thompson R.G. , and Thrall R.M. 1996. "Extensions and New Developments in DEA." *Annals of Operations Research* 66:3-45.
71. De Borger B. , Ferrier G. D. , and Kerstens K. 1998. "The Choice of a Technical Efficiency Measure on the Free Disposal Hull Technology: A Comparison Using US Banking Data." *European Journal of Operational Research* 105:427-46.
72. Debreu G. 1951. "The Coefficient of Resource Utilisation." *Econometrica* 19(3):273-92.
73. Deprins D. , Simar L. , and Tulkens H. 1984. "Measuring Labor-Efficiency in Post Offices." Pp. 243-67 in *The Performance of Public Enterprises: Concepts and Measurement*, Marchand M. , Pestieau P. , and Tulkens H. Amsterdam: Elsevier.
74. Dermine J., ed. 1993. *European Banking in the 1990s*. Great Britain: Blackwell Business.
75. Doyle P. , Fenwick I. , and Savage G.P. 1981. "A Model for Evaluating Branch Location and Performance." *Journal of Bank Research* 12(2):90-95.
76. Drake L. and Howcroft B. 1994. "Relative Efficiency in the Branch Network of a U.K. Bank: An Empirical Study." *Omega, International Journal of Management Science* 22(1):83-90.
77. Drake L. and Weyman-Jones T. G. 1996. "Productive and Allocative Inefficiencies in U.K. Building Societies: a Comparison of Non-Parametric and Stochastic Frontier Techniques." *The Manchester School* 64(1):22-37.
78. Dyson R. G. and Thanassoulis E. 1988. "Reducing Weight Flexibility in Data Envelopment Analysis." *Journal of Operational Research Society* 38(6):563-76.
79. Dyson R. G. , Thanassoulis E. , and Boussofiane A. 1990. "Data Envelopment Analysis." Pp. 13-28 in *Operational Research Tutorial Papers*, Hendry L.C. and Eglese R. The Operational Research Society.
80. Efron B. and Tibshirani R.J. 1993. "An Introduction to the Bootstrap." Monographs on Statistics and Applied Probability, No. 57, Chapman and Hall, New York, London.
81. Elyasiani E. and Mehdiian S. 1995. "The Comparative Efficiency Performance of Small and Large US Commercial Banks in the Pre- and Post-Deregulation Eras." *Applied Economics* 27:1069-79.

82. Elyasiani E. , Mehdian S. , and Rezvanian R. 1994. "An Empirical Test of the Association Between Production and Financial Performance: the Case of the Commercial Banking Industry." *Applied Financial Economics* 4:55-59.
83. Fare R. , Grifell-Tatjé E. , Grosskopf S. , and Lovell C.A.K. 1997. "Biased Technical Change and the Malmquist Index." *Scandinavian Journal of Economics* 99(1):119-27.
84. Fare R. and Grosskopf S. 1996. *Intertemporal Production Frontiers: With Dynamic DEA*. Kluwer Academic Publishers.
85. Fare R. , Grosskopf S. , Lindgren B. , and Roos P. 1994. "Productivity Developments in Swedish Hospitals: A Malmquist Output Index Approach." Pp. 253-72 in *Data Envelopment Analysis: Theory, Methodology and Applications*, Charnes A. , Cooper W. W. , Lewin A. Y. , and Seiford L. M. Boston: Klumer Academic Publishers.
86. Fare R. , Grosskopf S. , and Lovell C. A. K. 1985. *The Measurement of Efficiency of Production*. Klumer-Nijhoff Publishing.
87. Fare R. , Grosskopf S. , and Lovell C. A.K. 1994. *Production Frontiers*. Cambridge University Press.
88. Fare R. , Grosskopf S. , Lovell C. A. K. , and Yaisawarng S. 1993. "Derivation of Shadow Prices for Undesirable Outputs: a Distance Function Approach." *Review of Economics and Statistics* 75(2):374-80.
89. Fare R., Grosskopf S. , and Roos P. 1998. "Malmquist Productivity Indexes: a Survey of Theory and Practice." Pp. 127-90 in *Index Numbers: Essays in Honour of Sten Malmquist*, Fare R. , Grosskopf S. , and Russell R.R. (eds). Kluwer Academic Publishers.
90. Fare R. , Grosskopf S. , and Weber W.L. 1997. "The Effect of Risk-Based Capital Requirements on Profit Efficiency in Banking." *Discussion Paper Series No. 97, Department of Economics, Southern Illinois University, Carbondale*.
91. Fare R. and Lovell C. A. K. 1978. "Measuring the Technical Efficiency of Production." *Journal of Economic Theory* 19:150-162.
92. Fare R. , Lovell C. A. K. , and Zieschang K. D. 1983. "Measuring the Technical Efficiency of Multiple Output Production Technologies." Pp. 159-71 in *Quantitative Studies on Production and Prices*, Eichhorn W. , Neumann K. , and Shephard R.W. Wurzburg: Physica-Verlag.
93. Farrell M. J. 1957. "The Measurement of Productive Efficiency." *Journal of the Royal Statistical Society Serie A*:252-67.
94. Favero C. A. and Papi L. 1995. "Technical Efficiency and Scale Efficiency in the Italian Banking Sector: a Non-Parametric Approach." *Applied Economics* 27:385-95.

95. Fernandez-Castro A. and Smith P. 1994. "Towards a General Non-Parametric Model of Corporate Performance." *Omega, International Journal of Management Science* 22(3):237-49.
96. Ferrier G. D. and Lovell C. A. K. 1990. "Measuring Cost Efficiency in Banking: Econometric and Linear Programming Evidence." *Journal of Econometrics* 46:229-45.
97. Fixler D.J. and Zieschang K.D. 1993. "An Index Number Approach to Measuring Bank Efficiency: An Application to Mergers." *Journal of Banking and Finance* 17:437-50.
98. Forsund F. R. , Lovell C.A.K. , and Schmidt P. 1980. "A Survey of Frontier Production Functions and of Their Relationship to Efficiency Measurement." *Journal of Econometrics* 13:5-25.
99. Fried H. O. , Lovell C. A. K. , and Schmidt S. S. 1993. *The Measurement of Productive Efficiency: Techniques and Applications*. Oxford University Press.
100. Fukuyama H. 1996. "Returns to Scale and Efficiency of Credit Associations in Japan: A Nonparametric Frontier Approach." *Japan and the World Economy* 8:259-77.
101. ———. 1993. "Technical and Scale Efficiency of Japanese Commercial Banks: a Non-Parametric Approach." *Journal of Econometrics* 25:1101-12.
102. Giokas D. 1991. "Bank Branch Operating Efficiency: A Comparative Application of DEA and the Loglinear Model." *Omega, International Journal of Management Science* 19(6):549-57.
103. Golany B. and Roll Y. 1989. "An Application Procedure for DEA." *Omega, International Journal of Management Science* 17(3):237-50.
104. Gong B. and Sickles R. C. 1992. "Finite Sample Evidence on the Performance of Stochastic Frontiers and Data Envelopment Analysis Using Panel Data." *Journal of Econometrics* 51:259-84.
105. Grabowski R. , Rangan N. , and Rezvanian R. 1993. "Organisational Forms in Banking: An Empirical Investigation of Cost Efficiency." *Journal of Banking and Finance* 17:531-38.
106. Grifell-Tatjé E. and Lovell C.A.K. 1997. "A DEA-Based Analysis of Productivity Change and Intertemporal Managerial Performance." *Annals of Operations Research* 73(177-189.).
107. Grifell-Tatjé E. and Lovell C. A. K. 1996. "Deregulation and Productivity Decline: The Case of Spanish Savings Banks." *European Economic Review* 40:1281-303.
108. Grifell-Tatjé E. and Lovell C.A.K. 1995. "A Note on the Malmquist Productivity Index." *Economics Letters* 47:169-75.

109. Grifell-Tatjé E. and Lovell C. A. K. 1997. "The Sources of Productivity Change in Spanish Banking." *European Journal of Operational Research* 98:364-80.
110. Hancock D. 1985. "Bank Profitability, Interest Rates and Monetary Policy." *Journal of Money, Credit and Banking* 17(2):189-202.
111. ———. 1985. "The Financial Firm: Production With Monetary and Nonmonetary Goods." *Journal of Political Economy* 93(5):859-80.
112. Hartman T. and Storbeck J. E. 1996. "Input Congestion in Loan Operations." *International Journal Production Economics* 46-47:413-21.
113. Hermalin B. E. and Wallace N. E. 1994. "The Determinants of Efficiency and Solvency in Savings and Loans." *RAND Journal of Economics* 25(3):361-81.
114. Humphrey D. B. 1985. "Costs and Scale Economies in Bank Intermediation." *Handbook of Banking Strategy*, Aspinwall R. C. // Eisenbeis R. New York: Wiley.
115. Kao C. 1994. "Efficiency Improvement in Data Envelopment Analysis." *European Journal of Operational Research* 73:487-94.
116. Kaplan R. S. and Norton D. P. 1992. "The Balanced Scorecard - Measures That Drive Performance." *Harvard Business Review* January-February:71-79.
117. Kerstens K. and Vanden Eeckaut P. 1999. "Estimating Returns to Scale Using Non-Parametric Deterministic Technologies: A New Method Based on Goodness-of-Fit." *European Journal of Operational Research* 113:206-14.
118. Koopmans. 1957. *Activity Analysis of Production and Allocation*. New York: John Wiley and Sons.
119. Kuussaari H. 1993. "Productive Efficiency in Finnish Local Banking During 1985-1990." *Bank of Finland Discussion Paper* 14/93.
120. Linder J.C. and Crane D.B. 1992. "Bank Mergers: Integration and Profitability." *Journal of Financial Services Research* 7:35-55.
121. Lovell C. A. K. and Schmidt P. 1988. "A Comparison of Alternative Approaches to the Measurement of Productive Efficiency." Pp. 3-32 in *Applications of Modern Production Theory: Efficiency and Productivity*, Dogramaci A. and Fare R. (eds).
122. Malmquist S. 1953. "Index Numbers and Indifference Surfaces." *Trabajos De Estadística* 4:209-42.
123. McAllister P. and McManus D. 1993. "Resolving the Scale Efficiency Puzzle in Banking." *Journal of Banking and Finance* 17:389-405.
124. Miller S. M. and Noulas A. G. 1996. "The Technical Efficiency of Large Bank Production." *Journal of Banking and Finance* 20:495-509.

125. Molyneux P. , Altunbas Y. , and Gardener E. 1996. *Efficiency in European Banking*. Great Britain: Wiley.
126. Norton R. 1994. "Economics for Managers: Which Offices or Stores Really Perform Best? A New Tool Tells." *Fortune* October 31:26.
127. Nunamaker T. R. 1985. "Using Data Envelopment Analysis to Measure the Efficiency of Non-Profit Organisations: a Critical Evaluation." *Managerial and Decision Economics* 6(1):50-58.
128. Oral M. , Kettani O. , and Yolalan R. 1992. "An Empirical Study on Analysing the Productivity of Bank Branches." *IIE Transactions* 24(5):166-76.
129. Oral M. and Yolalan R. 1990. "An Empirical Study on Measuring Operating Efficiency and Profitability of Bank Branches." *European Journal of Operational Research* 46:282-94.
130. Parkan C. 1987. "Measuring the Efficiency of Service Operations: An Application to Bank Branches." *Engineering Costs and Production Economics* 12:237-42.
131. Pastor J. T. , Perez F. , and Quesada J. 1997. "Efficiency Analysis in Banking Firms: An International Comparison." *European Journal of Operational Research* 98:395-407.
132. Pastor J. T. , Ruiz J.L. , and Sirvent I. 1999. "An Enhanced DEA Russell Graph Efficiency Measure." *European Journal of Operational Research*:596-607.
133. Peristiani S. 1997. "Do Mergers Improve the X-Efficiency and Scale Efficiency of U.S. Banks? Evidence From the 1980s." *Journal of Money, Credit and Banking* 29(3):326-37.
134. Petersen N. C. 1990. "Data Envelopment Analysis on a Relaxed Set of Assumptions." *Management Science* 36(6):305-14.
135. Rangan N. , Grabowski R. , Aly H. Y. , and Pasurka C. 1988. "The Technical Efficiency of US Banks." *Economics Letters* 28:169-75.
136. Resti A. 1997. "Evaluating the Cost-Efficiency of the Italian Banking System: What Can Be Learned From the Joint Application of Parametric and Non-Parametric Techniques." *Journal of Banking and Finance* 21(2):221-50.
137. Revell J. (ed). 1994. *The Changing Face of European Banks and Securities Market*. United states: St. Martin's Press.
138. Rhoades S.A. 1993. "Efficiency Effects of Horizontal (in-Market) Bank Mergers." *Journal of Banking and Finance* 17:411-22.
139. ———. 1998. "The Efficiency Effects of Bank Mergers: An Overview of Case Studies of Nine Mergers." *Journal of Banking and Finance* 22:273-91.

140. Rouse P. , Putterill M. , and Ryan D. 1997. "Towards a General Framework for Performance Measurement: a Comprehensive Highway Maintenance Application." *The Journal of Productivity Analysis* 8(2):127-49.
141. Saunders A. 1993. "Panel Discussion of 'The Efficiency of Financial Institutions Around the Globe'." *Journal of Banking and Finance* 17((2-3)):551-57.
142. Schaffnit C. and Paradi J. C. 1998. "Performance Evaluation of the Commercial Branch Network of a Large Canadian Bank Using DEA." *Working Paper, Centre for Management of Technology and Entrepreneurship, Faculty of Applied Science and Engineering, University of Toronto*.
143. Schaffnit C. , Rosen D. , and Paradi J. C. 1997. "Best Practice Analysis of Bank Branches: An Application of DEA in a Large Canadian Bank." *European Journal of Operational Research* 98:269-89.
144. Schmidt P. 1985. "Frontier Production Functions." *Econometric Reviews* 4(2):289-328.
145. Sealey C. W. and Lindley J. T. 1977. "Inputs, Outputs and a Theory of Production and Cost at Depository Financial Institutions." *Journal of Finance* 32(4):1251-66.
146. Seiford L. M. 1996. "Data Envelopment Analysis: The Evolution of the State-of-the-Art (1978-1995)." *Journal of Productivity Analysis* 7(2-3):99-137.
147. Seiford L. M. and Thrall R. M. 1990. "Recent Developments in DEA." *Journal of Econometrics* 46:7-38.
148. Seiford L.M. and Zhu J. 1999. "An Investigation of Returns to Scale in Data Envelopment Analysis." *Omega, International Journal of Management Science* 27:1-11.
149. Sexton T. R. , Silkman R. H. , and Hogan A. J. 1986. "Data Envelopment Analysis: Critique and Extensions." *Measuring Efficiency: An Assessment of Data Envelopment Analysis*, Silkman R. H. New directions for program evaluation.
150. Sheldon G. 1994. "Economies, Inefficiencies and Technical Progress in Swiss Banking." Pp. 115-32 in *The Competitiveness of Financial Institutions and Centres in Europe*, vol. 28, Financial and Monetary Policy Studies, Fair D. and Raymond R. Dordrecht: Kluwer Academic Publishers.
151. Sheldon G. and Haegler U. 1993. "Economies of Scale and Scope and Inefficiencies in Swiss Banking." Pp. 103-40 in *Banking in Switzerland*, Blattner N. , Genberg H., and Swoboda A. (eds). New York: Springer-Verlag.
152. Shephard R.W. 1953. *Cost and Production Functions*. Princeton, NJ: Princeton University Press.

153. ———. 1970. *Theory of Cost and Production Functions*. Princeton University Press.
154. Sherman H. D. and Gold F. 1985. "Bank Branch Operating Efficiency: Evaluation With Data Envelopment Analysis." *Journal of Banking and Finance* 9:297-315.
155. Sherman H. D. and Ladino G. 1995. "Managing Bank Productivity Using Data Envelopment Analysis (DEA)." *Interfaces* 25(2):60-73.
156. Sherman H. D. and Rupert T. 1999. "Delays in Consolidation Can Be Costy." *American Banker*(March, 16th).
157. Siegel S. and Castellan N.J. 1988. *Nonparametric Statistics for the Behavioral Sciences*. Singapore: McGraw-Hill.
158. Simar L. and Wilson P. W. 1998. "Sensitivity Analysis of Efficiency Scores: How to Bootstrap in Nonparametric Models." *Management Science* 44(1):49-61.
159. Smith P. 1990. "Data Envelopment Analysis Applied to Financial Statements." *Omega, International Journal of Management Science* 18(2):131-38.
160. ———. 1997. "Model Misspecification in Data Envelopment Analysis." *Annals of Operations Research* 73:233-52.
161. Soenen L. A. 1974. "Locating Bank Branches." *Industrial Marketing Management* 3: 211-28.
162. Soteriou A.C. and Stavriniades Y. 1997. "An Internal Customer Service Quality Data Envelopment Analysis Model for Bank Branches." *International Journal of Operations and Production Management* 17(8):780-789.
163. Soteriou A.C. and Zenios S. A. 1999. "Using Data Envelopment Analysis for Costing Bank Products." *European Journal of Operational Research* 114:234-48.
164. Stolp C. 1990. "Strengths and Weaknesses of Data Envelopment Analysis: an Urban and Regional Perspective." *Computers, Environment and Urban Systems* 14:103-16.
165. Taylor W.M. , Thompson R. G. , Thrall R. M. , and Dharmapala P.S. 1997. "DEA/AR Efficiency and Profitability of Mexican Banks: A Total Income Model." *European Journal of Operational Research* 98:346-63.
166. Thanassoulis E. , Boussofiane A. , and Dyson R. G. 1996. "A Comparison of Data Envelopment Analysis and Ratio Analysis As Tools for Performance Assessment." *Omega, International Journal of Management Science* 24(3):229-44.
167. ———. 1995. "Exploring Output Quality Targets in the Provision of Perinatal Care in England Using Data Envelopment Analysis." *European Journal of Operational Research* 80:588-607.

168. Thanassoulis E. and Dunstan P. 1994. "Guiding Schools to Improved Performance Using Data Envelopment Analysis: An Illustration With Data From a Local Education Authority." *Journal of Operational Research Society* 45(11):1247-62.
169. Thanassoulis E. and Dyson R. G. 1992. "Estimating Preferred Target Input-Output Levels Using Data Envelopment Analysis." *European Journal of Operational Research* 56:80-97.
170. Thompson R. G. , Brinkmann E. J. , Dharmapala P. S. , Gonzalez-Lima M. D. , and Thrall R. M. 1997. "DEA/AR Profit Ratios and Sensitivity of 100 Large U.S. Banks." *European Journal of Operational Research* 98:213-29.
171. Thompson R. G. , Dharmapala P. S. , Humphrey D. B. , Taylor W. M. , and Thrall R. M. 1996. "Computing DEA/AR Efficiency and Profit Ratio Measures With an Illustrative Bank Application." *Annals of Operations Research* 68:303-27.
172. Thompson R. G. , Dharmapala P. S. , Rothenberg L. J. , and Thrall R. M. 1996. "DEA/AR Efficiency and Profitability of 14 Major Oil Companies in U.S. Exploration and Production." *Computers and Operations Research* 23(4):357-73.
173. Thompson R. G. , Langemeier L. N. , Lee C. , Lee E. , and Thrall R. M. 1990. "The Role of Multiplier Bounds in Efficiency Analysis With Application to Kansas Farming." *Journal of Econometrics* 46:93-108.
174. Thompson R. G. , Singleton F. D. , Thrall R. M. , and Smith B. A. 1986. "Comparative Site Evaluations for Locating a High-Energy Physics Lab in Texas." *Interfaces* 16(6):35-49.
175. Thompson R. G. and Thrall R. M. 1994. "Polyhedral Assurance Regions With Linked Constraints." Pp. 121-33 in *New Directions in Computational Economics*, W.W. Charnes and A. B. Whinston. New York: Kluwer Academic Publishers.
176. Tulkens H. 1993. "On FDH Efficiency Analysis: Some Methodological Issues and Applications to Retail Banking, Courts and Urban Transit." *The Journal of Productivity Analysis* 4:183-210.
177. Tulkens H. and Vanden Eeckaut P. 1995. "Non-Parametric Efficiency, Progress and Regress Measures for Panel Data: Methodological Aspects." *European Journal of Operational Research* 80:474-99.
178. Vassiloglou M. and Giokas D. 1990. "A Study of the Relative Efficiency of Bank Branches: An Application of Data Envelopment Analysis." *Journal of the Operational Research Society* 41(7):591-97.
179. Vennet R. V. 1996. "The Effect of Mergers and Aquisitions on the Efficiency and Profitability of EC Credit Institutions." *Journal of Banking and Finance* 20:1531-58.

180. Ward P.T. , Storbeck J.E. , Mangum S.L. , and Byrnes P. 1997. "An Analysis of Staffing Efficiency in U.S. Manufacturing: 1983 and 1989." *Annals of Operations Research* 73:67-89.
181. Wheelock D.C. and Wilson P. W. 1995. "Explaining Bank Failures: Deposit Insurance, Regulation, and Efficiency." *Review of Economics and Statistics* 77(4):689-700.
182. Wong Y.-H. B. and Beasley J. E. 1990. "Restricting Weight Flexibility in Data Envelopment Analysis." *Journal of the Operational Research Society* 41(9):829-35.
183. Yeh Q. 1996. "The Application of Data Envelopment Analysis in Conjunction With Financial Ratios for Bank Performance Evaluation." *Journal of the Operational Research Society* 47:980-988.
184. Zaim O. 1995. "The Effect of Financial Liberalization on the Efficiency of Turkish Commercial Banks." *Applied Financial Economics* 5:257-64.
185. Zhu J. 1996. "Data Envelopment Analysis With Preference Structure." *Journal of the Operational Research Society* 47(1):136-50.
186. Zieschang K. D. 1984. "An Extended Farrell Technical Efficiency Measure." *Journal of Economic Theory* 33:387-96.